School-Track Environment or Endowment: What Determines Different Other-Regarding Behavior Across Peer Groups?*

Katrin John†
Otto-von-Guericke-University of Magdeburg, Leibniz University Hannover & NIW Hannover

Stephan L. Thomsen‡
NIW Hannover, Leibniz University Hannover & ZEW Mannheim

This version: October 24, 2013

Abstract

We demonstrate that school-track environment is a determinant in the formation of other-regarding preferences. Using experimental data of German schoolchildren, we compare other-regarding behavior across two school tracks resulting from achievement grouping. Across the tracks, pupils differ in their share of giving in the dictator game as well as in the shares of free-riders and social welfare maximizers. Although tracking also results in deviant peer-groups regarding individual endowment, differences in IQ scores and self-evaluated personality cannot explain the gap in experimental outcomes. A matching analysis confirms that there is a robust treatment effect of school-track environment on other-regarding preferences.

Keywords: Other-regarding preferences, Tracking, Dictator game, Public goods game, Matching
JEL Classification: C91, D03, I21

*We thank Dirk Sliwka for helpful comments as well as discussants at the GfeW conference 2013. Financial support from the Wissenschaftszentrum Sachsen-Anhalt Lutherstadt Wittenberg (WZW) in the course of the project "Analyse des Bestands und der ökonomischen Bedeutung kognitiver und nicht-kognitiver Fähigkeiten in Sachsen-Anhalt zur Identifikation (bildungs-)politischer Handlungsbedarfe" is gratefully acknowledged.

†Katrin John is Research Assistant at Otto-von-Guericke-University Magdeburg, Department of Economics and Management, and working at NIW, Hannover. Address: NIW, Königstraße 53, D-30175 Hannover, e-mail: john@niw.de, phone: +49 511 12331636.

‡Stephan L. Thomsen is Director of NIW, Hannover, Professor of Applied Economics at Leibniz University Hannover, and Research Associate at ZEW, Mannheim. Address: NIW, Königstraße 53, D-30175 Hannover, e-mail: thomsen@niw.de, phone: +49 511 12331632, fax: +49 511 12331655.

1 Introduction

Childhood and adolescence constitute a sensitive period for the formation of other-regarding preferences. As children grow older, they exhibit more other-regarding behavior as revealed by their choices in standard economic games (see Murninghan and Saxon, 1998; Harbaugh and Krause, 2000; Fehr et al., 2008; or Martinsson et al., 2011). Since there is compulsory schooling during most of this period, school environment is an obvious object of investigation in view of the process of preference formation. During their school time, pupils are subject to tracking, that is grouping by ability. This process can be regarded as a mechanism that creates specific environments within the tracks. School tracking has been shown to result in particular peer groups not only with respect to ability but also regarding socio-economic background (see Dustmann, 2004, or Tamm, 2008). In light of the existing evidence on peer effects for various outcomes ranging from achievement (see Betts, 2011, for an overview) to substance use (see, e. g., Kremer and Levy, 2008), school tracking therefore provides an interesting instrument of how environmental effects influence the formation of other-regarding preferences.

In the course of a series of behavioral games played with German school children from two different school tracks, John and Thomsen (2013) show, that behavior in dictator and public goods games not only varies across age groups, but also with respect to culture and school track. Thereby, differences with respect to school track were as large as those between age groups. Distinguishing between two school tracks – the lower, non-academic track and the higher, academic track, they show that in the dictator game, pupils from the higher track on average gave more to their partner than pupils in the lower track. In the public goods game, however, pupils from the lower track exhibited a lower share of free-riders as well as a higher share of social welfare optimizers compared to pupils in the higher track. These results are taken as a starting point to the analysis at hand.

We will pursue two explanations for the observed difference in behavior. One possible influence is that peer groups may operate through the creation of track-specific behavioral norms. Especially during adolescence, peer groups constitute an important source for behavioral orientation (see, e. g., Meeus et al., 2002). It is therefore likely, that peer groups emerging from tracking set up track-specific behavioral norms with respect to situations where other-regarding preferences matter. This explanation also relates to the concept of a social identity (see Akerlof and Kranton, 2000) with identity-specific behavioral norms. Akerlof and Kranton (2002) depict an economic model of social identity in schools. The model states that utility is highest when a pupil's characteristics and behavior match the ideal of the social category he or she has chosen. Hence, while peer-models focus on the personal characteristics and the background, identity economics additionally emphasizes norms and prescriptions that evolve from the group-specific situation or environment. Of course, both explanations can apply simultaneously.

While it is straightforward to suggest school-track environment as a driver of differences in other-regarding preferences, track-specific environments are the outcome of self-selection with respect to achievement and possibly other characteristics. This leads to pupils differing in their

individual characteristics due to tracking. In case that individual characteristics contribute to the formation of other-regarding preferences, differences in experimental outcomes across school tracks may result from the self-selection process. Therefore, the second possible explanation is that differences in individual characteristics account for the behavior. This explanation would relate to a peer composition effect on other-regarding preferences.

Using IQ scores and self-evaluated personality traits as measures of individual characteristics, we test the second hypothesis, namely that differences in endowment contribute to the differences in outcomes. Applying several checks on the experimental outcomes, we can establish differences in outcomes with respect to some of the individual characteristics. For instance, our data reveal a link between less socially desirable personality profiles and altruistic behavior. Nevertheless, our results mainly support the first explanation, namely that school-track environment causes differences in other-regarding preferences. We perform a matching analysis to account for the existing differences in individual characteristics across tracks. Comparing similar individuals across tracks enables a robust estimate of the actual influence of school-track environment on other-regarding preferences while controlling for self-selection. Results of the matching analysis show, that track differences remain significant even when we condition the comparison on individual characteristics. In view of our research question, we therefore conclude, that tracking contributes to the environmental influence on other-regarding preferences. This implies another dimension added to the literature on tracking. Besides aspects of achievement, the process of tracking leads to different economic preferences.

The course of the paper is as follows: section 2 is a summary on school tracking and its implementation. Section 3 presents a model of tracking and other-regarding preferences. Information on the experimental design as well as on different aspects of the data is provided in section 4. Section 5 encompasses all results from univariate analysis, the matching analysis as well as an additional robustness check. Finally, section 6 discusses possible limitations and implications of our research and section 7 concludes.

2 The Institutional Setting of School Tracking

School tracking denotes achievement grouping of students with the aim to adjust teaching to the distinct needs of groups differing by ability.[1] Within the groups, adjusted teaching can be implemented with the help of different curricula, class sizes, teaching styles, adjustment of instruction time, or other resources related to the school environment. There are two organizational forms of tracking: grouping students in different classrooms at the same school, or streaming students in different schools. Tracking students into different schools usually implies an academic and one or more non-academic tracks with specific curricula, adjusted to whether there will be academic post-secondary education or vocational education after finishing the respective school. This form of tracking is widespread in European countries, as for instance in Germany, Italy,

[1] This summary follows Betts (2011) introduction to the economics of tracking.

Austria, Belgium, Czech Republic, Hungary, Ireland, Luxembourg, the Netherlands, Portugal, Slovak Republic, Slovenia, Switzerland and others (OECD, 2012).[2]

In Germany, school tracking is implemented at the age of 10 and there are three different school types available.[3] The lowest school track is the *Hauptschule* (lower secondary school track) where a basic general education is taught until grade 9 (about age 15 or 16). The leaving certificate is the prerequisite for starting vocational training. Some forms of training (mostly white-collar professions) require the leaving certificate earned at the *Realschule* (upper secondary school track), where a more extensive general education is taught until grade 10 (about age 16 or 17). At the *Gymnasium* (grammar school track), pupils earn an intensified general education up to grade 12 or 13 (about age 18 or 19). The school leaving certificate of the grammar school enables pupils to go to university. During their last year of primary school, pupils are given a recommendation as to which school form is most appropriate for their secondary education, based on a pupil's marks and behavior. Whether this recommendation is binding depends on the law of the federal state. In the states where we ran the experiments, it is non-binding, leaving parents to decide about their children's school track. In Germany, about 39% of all pupils attended grammar schools in 2011.[4]

Regarding the aim of this paper, there are two important aspects of school tracking, that need to be considered in detail. The first relates to the determinants of tracking. As mentioned above, achievement is the central aspect of the tracking decision. Nevertheless, Dustmann (2004), Mühlenweg and Puhani (2010) and Tamm (2008) show for Germany, that other aspects such as the socio-economic background, birth date, and motivation or expectation of parents have an impact on the actual placement within tracks. Hindriks et al. (2010) present similar results for Belgium, where the tracking system largely resembles that of Germany. In order to detect in how far differences in other-regarding preferences are due to the school-track environment, we have to account for differences in endowment caused by the tracking procedure.

The other aspect is the creation of track-specific conditions when streaming into separate schools is in place. As Betts (2011) points out, this usually implies different resources allocated at the different school types. Pupil-teacher ratios differ across tracks leading to an adaptation of instruction time and related pedagogical methods. In addition, tracking creates different peer groups (see also Betts, 2011). While within-school tracking leaves the school-level peer-group unchanged, streaming pupils into different schools creates locally distinct peer-groups. Conditional on the determinants of tracking, this results in differences with respect to ability and socio-economic background as the most important influences. When studying the possible

[2]Tracking pupils in different classrooms at the same school is the common form of tracking in Australia, Canada and the US. Within Europe, the Scandinavian countries, Poland and Spain, for instance, follow the comprehensive school system with tracking starting only after the age of 15.

[3]There are a number of additional school types combining two or three streams. Since educational legislative is situated at the federal level, the availability of these school types differs across the Laender. As the degrees earned at these schools do not differ from those at schools with one single track, we will not pursue the description of these institutions.

[4]This figure relates to the share of pupils leaving elementary school and pursuing their secondary education at a grammar school. Source: Federal Statistical Office Germany (2012a).

impact of school tracking on other-regarding preferences, the aspects just described represent potential transmission mechanisms.

3 A Model of Tracking and Other-Regarding Preferences

We apply the Roy model to investigate the process of tracking and its possible impact on other-regarding preferences. For the sake of simplicity we distinguish two possible tracks – a low and a high track. Relating to the description in section 2, these can be thought of as a non-academic and an academic track in the case of streaming pupils. Similarly, the model can be applied to the case of within-school tracking, where lower and higher achievement classes can be discriminated.

The tracking decision is linked to individual characteristics. Hence, we generally refer to sorting models with a wider applicability such as the labor market. While ability is the determinant of tracking in all countries, other aspects, such as family background, can be incorporated as well. After the tracking decision, individuals will find themselves in a track-specific environment with the resulting peer-group. This environment implies a particular socialization which brings along particular beliefs and expectations. Following the approach of Akerlof and Kranton (2000, 2002), individuals choose to adopt identities, which have "prescriptions" on how to behave in situations. These prescriptions may include norms on how to act in group activities such as the public goods' task, that we implemented. And depending on the identity, that is related to a school track, prescriptions can differ across tracks. Of course, we are not able to empirically distinguish between the concepts of environment and identity.

The main focus of our analysis are other-regarding preferences. Denote other-regarding preferences y as $y = \mu(\mathbf{X})$, being a function of some individual and environmental characteristics $\mathbf{X}$. For the sake of simplicity, assume other-regarding preferences to be represented by a scalar, where y_i denotes the individual value. Based on the outcome of interest, this scalar can be interpreted as the degree of altruism, or as the disposition to free-ride, or choose the social optimum. Now, we think of tracking constituting one of the possible influences of y, that is, being part of the characteristics $\mathbf{X}$. Then, more formally, we can describe the model analogously to a treatment model with track-specific environment as the treatment on the other-regarding preference outcome. In reference to the returns of schooling we thus examine whether school track has an effect on other-regarding preferences.

Since the tracking decision is endogenous, the potential outcome model proposed by Roy (1951) and formalized by Rubin (1974) provides an appropriate framework for the analysis. Denote school track by the indicator D, with $D = 1$ when being in higher track school. Analogously to the treatment model, the school track assignment can be represented by a latent variable D^*, that is a function of some variables $\mathbf{Z}$, where $\mathbf{X}$ may be a subset of $\mathbf{Z}$. Hereby, individual characteristics are linked to the recommendation of school track, that is we have an endogenous treatment. The participation equation links D^* to D. As explained above, we observe pupils

in different school tracks, that is after the assignment of treatment has taken place. Hence, we observe individuals' other-regarding preferences in either one of the tracks, that is $(y|D=0)$ or $(y|D=1)$, our potential outcomes. The model can then be summarized as below.

$$y_0 = \mu_0(\mathbf{X}, u_0) \tag{1}$$

$$y_1 = \mu_1(\mathbf{X}, u_1) \tag{2}$$

$$D^* = \mu_D(\mathbf{Z}, u_Z) \tag{3}$$

$$D = \begin{cases} 1 & D^* > 0 \\ 0 & D^* \leq 0 \end{cases} \tag{4}$$

Since tracking is related to characteristics $\mathbf{X}$ with variables from $\mathbf{X}$ being in $\mathbf{Z}$, we have $corr(\mathbf{X}, D) \neq 0$. Due to this correlation, examination of $E[y|\mathbf{X}]$ needs to account for the selection into the higher school track (treatment), as well as an analysis of $E[y|D]$ needs to account for the different distribution of $\mathbf{X}$ across tracks.

Besides the investigation of different conditional means within school-track environments, we would like to test whether school-track environment itself influences other-regarding preferences. This would be the case if $E[y|\mathbf{X}, D=1] \neq E[y|\mathbf{X}, D=0]$, conditioning on the same $\mathbf{X}$, which amounts to calculating the average treatment effect $E[y_1 - y_0] = \mu_1(\mathbf{X}) - \mu_0(\mathbf{X})$. This parameter refers to the expected effect a switch from the lower to the higher school track would have on the measure of other-regarding preferences.

Regarding the starting point of deviating other-regarding preferences across school tracks, we put forward two possible hypotheses: either, differences are due to different distributions of the $\mathbf{X}$. In this case, $E[y|\mathbf{X_1}, D=1] \neq E[y|\mathbf{X_0}, D=1]$. That is, our first hypothesis relates to a composition effect driving the difference in outcomes. Secondly, differences may be due to the different environments provided by the tracks. In this case, we should still observe differences in other-regarding preferences, when we compare matched individuals from both tracks, $E[y|\mathbf{X_1}, D=1] \neq E[y|\mathbf{X_1}, D=0]$. That is, our second hypothesis is the presence of a significant treatment effect.

4 Data Set and Experiment Design

4.1 Data Set

Our data set includes experimental data of German school children aged 10 to 18 years enrolled in four public schools in Germany, two in East Germany (federal state of Saxony-Anhalt) and two in West Germany (federal state of Lower-Saxony). The two federal states are neighboring (see Figure 1) with an approximate distance of 150 km between the schools in the respective other federal state, that is, despite the federal border regions are similar. In both federal states, we chose a secondary school and a grammar school. In Saxony-Anhalt, secondary schools are

schools with tracks for lower and upper secondary school forms. The school we chose had one class for the lower secondary track and two classes for the upper secondary track in grade 9. In the other grades there was no additional tracking. In Lower-Saxony, separate schools for the two secondary tracks exist, so we chose an upper secondary school. We will refer to lower and upper secondary schools as the lower track since both represent a non-academic track, while grammar schools with their academic curriculum will be denoted as the higher track. In the lower track schools, we covered pupils from grades 5, 7 and 9, while in the higher track schools, we also surveyed grade 11. Pupils in our sample study together in classes of between 19 and 32 students.

[Include Figure 1 here]

We collected data for 530 pupils for both games distributed across the two school tracks. There are 294 pupils in the lower track and 236 in the higher track. The split between lower and higher track is close to the actual division: in our sample, 46% of the pupils attend the higher track school, which is a little higher than the nationwide share of 39% in the school year 2011/2012. In the state of Lower Saxony 42% of the children followed the higher track in the school year 2011/2012, in our sample, 43% of the pupils surveyed there are in the higher track. For Saxony-Anhalt, our share of 45,7% is also very close to the actual share of 44.5% (Source: Federal Statistical Office Germany, 2012a).

Relying on pupils for the study of other-regarding preferences has the advantage, that we do not face distortions in the measurement of other-regarding preferences as can be the case for university student samples (see Anderson et al., 2012). Pupils are highly unlikely to have been introduced to economic game theory. Therefore, the risk of informed strategic behavior is reduced. Another advantage is, that our sample is non-selective with respect to incentives. Most samples to study other-regarding preferences rely on university students, who are likely to self-select due to monetary incentives (see Abeler and Nosenzo, 2013). Pupils in our sample where unaware of the experimental procedure as well as the incentives until right before participation.

The four schools where we surveyed the pupils have a very similar institutional setting. Therefore aggregation of the data across federal states is valid. Comparability is ensured by the fact, that all schools are public and non-specialized. In addition, they have an attendance zone which included the surrounding rural areas. That means, neighborhood biases arising from a specific peer-group in the quarter where the school is located are unlikely to occur. Furthermore, the respective communities resemble each other with respect to size, unemployment rate, and the share of employees subject to social insurance contribution.[5]

[5]Three of the four place are classified as semi-urban with 24,000 to 53,000 inhabitants, the fourth is a city with about 100,000 inhabitants. Unemployment rates only differ marginally, with three places having a rate of 6 and one place with 4%. The share of employees subject to social insurance contribution also shows little variation, ranging from 30 to 35%. (Source: Federal Statistical Office Germany, 2012d, for the population data, Federal Statistical Office Germany, 2012c, for the number of employees, and German Federal Employment Agency, 2012, for the number of unemployed.

4.2 Experiment Design

We played the dictator game (DG) and the public goods game (PGG). In each of the games, the endowment consisted of an envelope with 10 green plastic chips (tokens). Children had to decide upon the division of the endowment in each of the games. They were informed that each of the tokens had a monetary value and that at the end of each game, we would exchange their tokens for money that they could keep. The value depended on the grade: 5th graders received 0.10 EUR for one token, 7th graders 0.15 EUR, 9th graders 0.30 EUR, and 11th graders 0.50 EUR. The money at stake was chosen to represent the average weekly pocket money in each grade. On average, 7th graders earned €5, 9th graders €10.

The dictator games were administered as follows: half of the pupils in the experimental session were given two envelopes, one containing 10 tokens and another empty one. They were provided with the respective instructions and then put the amount that they wished to contribute into the empty envelope. We had prepared random assignments of pupils and distributed the envelopes accordingly. At the end of each game, we exchanged the tokens for cash. The game was played twice: those who were dictator in the first round were passive in the second round, but with a new random assignment. Children were explicitly told that assignments were changing in each round to prevent decisions being influenced by attempting to build up a reputation.

For the public goods game, each participant received two envelopes, one filled with the endowment of 10 tokens and another empty one for possible contributions to the public good. After we collected the contributions, children could turn around and observe us counting the contributed sum. We then doubled the contributions and publicly announced the individual share that each of the participants would receive in addition to their tokens possibly retained.

Moreover, the sequence of games varied randomly across sessions. Experiments were conducted by the same persons at all schools, ensuring that experimental conditions were comparable. There were three parallel sessions of 45 minutes each, all three together covering the pupils from one grade. On each day, there were four subsequent sessions at the grammar schools and three subsequent sessions at the secondary schools. Each of the sessions was headed by a researcher and supported by a student and sessions took place in classrooms with individual seating so that decisions could be made in private.

4.3 Intelligence and Personality Measures

Before the experimental sessions, all pupils completed an intelligence test, a questionnaire on personality and a background questionnaire, that covered questions on pupils' socio-demographic characteristics such as gender, date of birth, and the kind of household they live in. This procedure, as well as the subsequent experiments, was the same for all participants. Below we describe the measures for the endowment variables in detail since they are an essential part in testing the two hypotheses aiming to explain track-specific differences in other-regarding preferences.

4.3.1 Intelligence Measure

To elicit a measure for intelligence we used the revised Culture Fair Intelligence Test (CFT 20R), see Weiß (2006). It is a test to assess the general mental capacity ("g-factor") while avoiding culture-related influences ("culture fair"). In the context of our analysis of environment, this test is very appropriate, since it is a language-free test. This enables us to elicit a measure of intelligence, that is (mostly) independent of environmental influences of the subjects. Language-dependent intelligence tests, in contrast, may measure some kind of intelligence that is due to a specific environment, which is depicted by language (see Weiß, 2006). Test accomplishment took approximately 45 minutes including instructions. This was the first part executed and took place in the first school hour, so that pupils were still able to concentrate well. The CFT 20R encompasses four sub-tests, each using perceptual forms (continuing series, classification, matrices, topologies). In total, there were 56 items tested.

4.3.2 Personality Measures

To measure personality characteristics we used the Junior Temperament and Character Inventory (JTCI), see Goth and Schmeck (2009). It is based on the bio-psychological personality model of Cloninger (1994), that distinguishes four aspects of temperament and three aspects of character. Thereby, temperament traits are seen as quasi-automatic emotional reaction tendencies and habits in response to stimuli. Character traits, in contrast, are seen as expressing individual aims and values, which also depend on socio-cultural learning processes. Table 1 summarizes the seven traits, which are described below.

[Include Table 1 here]

The temperament part of personality is described by the dimensions novelty seeking, harm avoidance, reward dependence and persistence. Novelty seeking relates to the behavioral activation system and describes the extent of exploratory activity, as well as the approach to novel stimuli, and the avoidance of averse stimuli. Children and adolescents high in novelty seeking can be described as being impulsive, exploratory, or curious while those low in novelty seeking are characterized by being reflective and orderly.[6] Harm avoidance relates to the behavioral inhibition system and describes the extent of being shy, fatigable, fearful, or pessimistic. A high value of harm avoidance goes along with being shy, worrying, and pessimistic. Children and adolescents who are low in harm avoidance instead are confident, outgoing, and optimistic. Reward dependence relates to the behavioral dependence system and describes the extent of regulating one's behavior due to social cues. Those who are very reward dependent can be characterized as being attached, sentimental, and warm, while those not very reward dependent are described as detached, withdrawn, and cold. Finally, persistence relates to the behavioral

[6]The descriptions are taken from Cloninger (1994) and adjectives describing the traits are cited from Schmeck et al. (2006), Table 1.

persistence system which describes the intrinsic maintenance of a behavior. This temperament trait is associated with being either hard-working and ambitious, or on the other hand, with giving up easily and being inactive.

Character comprises the three dimensions self-directedness, cooperativeness, and self-transcendence. Self-directedness relates to the self-concept of an individual and describes responsibility and resourcefulness in order to achieve personal goals. A high disposition of this character trait refers to responsible and reliable individuals, whereas those low in self-directedness are immature and unreliable. Cooperativeness measures the extent of feeling as part of the social system. It comprises social tolerance, empathy, helpfulness, compassion and moral principles. Children and adolescents high in cooperativeness are characterized as being empathic, tolerant and helpful. On the contrary, individuals who have a low disposition of cooperativeness are revengeful and intolerant. Self-transcendence refers to the extent of feeling as part of one's surroundings as a unitive whole. It also encompasses the tendency of creative inspiration and spirituality. Having a high value of this character trait means being patient, creative and self-forgetful, whereas being low in self-transcendence is associated with being impatient and self-conscious. The individual dimensions of temperament and character are described as "causally independent but functionally interactive" (Cloninger, 1994[p. 267]).

Implementation of Personality Measures

To reduce dimensionality when evaluating social preferences by individual personality traits, we construct a variable describing the social desirability of an individuals' personality. All seven traits can be described as a continuum between two extremes subsumed by the adjectives in Table 1. One of these extremes is more socially desirable. Take for example cooperativeness: here, being empathic, tolerant, and helpful is socially higher valued than being revengeful and intolerant. Likewise, the following poles of the traits describe the "better" disposition: low values of novelty seeking and harm avoidance are preferred to high values, whereas for the other five traits high values indicate a more desirable personality aspect. Regarding that, in a first step we transform trait measures as deviations from the population median of 50. Thereby, the differences to the population median are always built such that a negative value indicates a less socially desirable value of the trait. Imagine an individual has a very low value with respect to cooperativeness, then we get a negative value of the transformed cooperativeness scale. On the opposite, a very high disposition of cooperativeness would lead to a positive value of the transformed scale.

Summing the deviations of all traits then provides a continuous measure of the overall desirability of an individual's personality, with higher values indicating a more socially desirable personality.[7] In order to test, whether other-regarding preferences vary with the level of social desirability of personality traits, we further define indicators related to the distribution of the

[7]The self-transcendence character trait is left-skewed in the sample. While the other traits have a median which is equal or almost identical to the population median of 50, the median for self-transcendence is equal to the 38 quantile and therefore far from the population median of 50. Since the distribution is the same across regions and school tracks, we leave it out of the calculation to prevent distortion of our variable of the sum of deviations from the medians.

variable just described. The first, denoted "personality score low", is equal to one, if a person's sum of deviations for all traits is located in the bottom 30% of the distribution. Figure 2 shows the kernel density estimate of the distribution of the sum of deviations from median traits.

[Include Figure 2 here]

The bottom 30% of the distribution are marked at the left of the figure. Individuals within this part of the distribution have an indicator equal to one with respect to the variable "personality score low". Another variable, denoted "personality score medium", indicates that the sum of deviations lies in the middle 40% of the distribution. Finally, an indicator denoted "personality score high" refers to the sum of deviations of all traits lying in the upper 30% of the distribution (right part of the kernel in Figure 2). Note that the low personality score indicates the least socially desirable personality type, that is, it refers to negative values of the deviations variable.

Besides the reduction in dimensionality, constructing indicators of social desirability of an individual's personality is meaningful with respect to our analysis. Other-regarding preferences are linked to norms and values present in a society. The compliance with existing norms and values is among other things related to personality (see, e. g., Austin et al., 2012). Often, pro-social behavior is also the socially appropriate behavior. And it has been shown for children, that more extreme types of personality are those who show less socially appropriate behavior (Eisenberg et al., 2006). Moreover, children with "difficult" temperament report behavioral problems more often (see Caspi, 1998). Since behavioral problems are linked to less pro-social behavior (Duncan et al., 1996), this also provides a link between personality types and other-regarding preferences. Therefore, having indicators for the degree of social desirability of personality allows to analyze this indirect relationship of personality and other-regarding preferences.

5 Results

5.1 Experimental Outcomes for Subgroups

To analyze how other-regarding preferences vary across different subgroups and environments, we provide summary tables. For both games, we display means of experimental outcomes with respect to school track, quartiles of the IQ distribution, as well as for the three broad types of personality (low, middle, and high social desirability). For the dictator game, Table 2 summarizes the results. Figures in bold face (italics) denote that differences between subgroups are significant at the 5%-level (10%-level) or lower for a Mann-Whitney-test.

[Include Table 2 here]

Mean giving in the dictator game amounted to 2.2 out of 10 tokens. Here, almost every fifth pupil (19%) kept all the tokens to himself or herself, while 8% chose an equal distribution.

Regarding school tracks, pupils from the higher track on average give more to their powerless partner in the dictator game. With respect to subgroups of the IQ distribution, the only significant difference occurs for pupils within the highest quartile of the distribution. They, on average, more often decide to keep all the endowment and share nothing. Further comparing school tracks within this group, there is no significant difference between tracks ($p = .17$ for a Mann-Whitney test). The finding that more intelligent pupils on average less frequently give a positive amount does not explain why higher track pupils on average give more in the dictator game.

For the different personality types, we find significantly different results for the least socially desirable type: pupils with a personality profile within the lowest 30% of the distribution of deviations from trait medians, on average give less than others, more frequently choose the egoistic outcome, and less frequently choose the equal split. Those, whose personality trait deviations from the medians are within the center 40% of the distribution more often chose the equal split, have a higher mean contribution, and the lowest share of choosing zero giving.

[Include Table 3 here]

In the public goods game (see Table 3), the average contribution to the public good amounted to 4.4 out of 10 tokens, with 8% of the pupils free-riding and 12% fully contributing their endowment. We do not find differences in the mean contribution for any of the subgroups examined, except for a marginally significantly higher mean giving for pupils within the lowest quartile of the IQ distribution. Nevertheless, there are differences in the extremes – zero and full contribution – for specific subgroups. Pupils from the higher track schools significantly less often made both extreme choices. Those with IQ scores in the lowest quartile of the distribution more often chose the socially efficient option in the public goods game and have a higher mean contribution. Pupils with the least socially desirable personality type exhibit the highest extent of social welfare preferences with only 5% of them free-riding, but 17% choosing the socially optimal amount. At the other end of the distribution, the share of choosing the socially optimal amount is lowest for pupils with a highly socially desirable personality type.

5.2 Endowment and its Relation to Experimental Outcomes

For endowment being a possible driver of the difference in outcomes across school tracks, we first need to ensure that the distribution of individual characteristics varies with the institutional environment. As indicated in the model above, individual characteristics, such as intelligence and personality influence the tracking decision. Hence, we would expect these variables to differ across school tracks. Table 4 shows means of these characteristics by school tracks. Mean IQ scores display sorting by ability due to the tracking decision. Moreover, all personality traits except for self-transcendence differ significantly in means across tracks. Figure 3 shows kernel density estimates of the distribution for all seven traits by school track. Living with a single

parent is used as an indicator for family background and shows considerably different values across tracks.

[Include Table 4 here]

[Include Figure 3 here]

We further examine the relationship of the endowment variables and measure of other-regarding preferences. This is of interest, because the direction of the relationship can indicate whether differences in endowment contribute to explain the difference in outcomes, hence produce a composition effect. Therefore, we show Spearman's rank correlation coefficients for IQ as well as for each of the personality traits and experimental outcomes. This is done for the overall sample as well as for the two school tracks, for IQ scores as well as for all seven personality traits. Table 5 indicates the correlation coefficients that are significant at the 5%-level.

[Include Table 5 here]

There are three noticeable aspects of the displayed correlation coefficients. First, there is only one relationship where we observe a significant coefficient within the overall sample as well as in both subsamples (public good contribution and persistence). Second, there are a number of relationships where a significant overall correlation coefficient is only replicated in the sample of the higher school track. There are, however, no reverse cases where it is only replicated in the lower track sample. And third, there are several relationships where we only observe a significant coefficient in the overall sample. With one exception, the second and third cases relate to insignificant coefficient(s) with the same sign but lower in magnitude. We therefore expect the lower number of observations in the subsamples to cause the pattern observed.

As regards content, we observe track-specific relations of IQ and experimental decisions: While higher IQ leads to more frequent egoistic choices in the DG for higher track pupils, it is related to lower public goods contributions for lower track pupils. Further, there is a negative relationship between novelty seeking and altruism. Individuals with a distinct disposition of novelty seeking (relating to being impulsive and agitable) are less likely to choose an equitable allocation in the dictator games and have a lower mean giving. Reward dependence displays a somewhat surprising pattern: while a high disposition of this trait (being sentimental, cordial) goes along with less frequent egoistic choices in the DG, this is also correlated with fewer social welfare maximizing choices in the PGG. It is important to bear in mind that it need not be the same pupils causing the two significant correlations, hence the two results need not be contradictory. A similar argument may hold for persistence, where a negative correlation with contributions to the public good go well along with a positive correlation with free-riding, but also with a negative correlation with the social welfare maximizing choice. In the same line, a high disposition of cooperativeness relates positively to dictator giving and egalitarian choices, but negatively to

choices of full contribution in the PGG. For all of the significant effects, correlation coefficients are between .09 and .21, hence not very high.

As measured by the number of significant correlations, cooperativeness and reward dependence seem to be the most important traits and they are the only traits, that are related to both games. While the relationship of cooperativeness and dictator giving points towards altruism as a driver of the results, at the same time high cooperativeness is also related to less frequent choice of the socially optimal amount. Similarly, reward dependence exhibits an unexpected relationship with the two games. While high disposition of this trait is linked to lower social welfare behavior, it is at the same time related to a less frequent choice to give nothing in the dictator game. It seems as if the public goods game, intended to measure social welfare preferences, is also subject to a maximization calculus. That is, self-concern is a motive of giving in the public goods game. Of course, there is an additional risk component of the returns to the public good compared to the secure return in the dictator game. Vesterlund (2006) summarizes evidence on self-interest as a motivation of why people contribute to the public good: When giving has a private benefit in form of warm-glow or avoidance of others' contempt, then we will observe positive contributions that, however, do not depict altruistic acts. This kind of explanation resolves the apparent contradiction of the sign of the relationship across the two games.

For both games we have noticed, that the more extreme subgroups (highest/lowest quartile of IQ distribution, least socially desirably personality type) are those for whom we observe deviations from the overall results (see Tables 2 and 3). In Table 6 we therefore provide a different editing of the results, evaluated for personality types and quartiles of the distribution of IQ scores at the same time. To keep the number of observations in the subsamples satisfactorily, we pooled the second and third quartile of the IQ-distribution as well as the first and the fourth one. Hence, we compare the "inner" half with the "outer" half of the distribution. The largest deviations from the overall results can be found for the "most extreme" subsample: pupils who have an IQ score in the outer half of the distribution and whose personality at the same time can be characterized as being least socially desirable exhibit the lowest mean giving in the dictator game as well as the highest mean contribution in the public goods game and the lowest share of free-riders. This group of children and adolescents amounts to one sixth of the complete sample and about two thirds are from the lower track school. And indeed, these results are not solely driven by pupils from the lowest 25% of the IQ distribution,[8] indicating that track differences are not driven by differences in individual characteristics. For the dictator game, we already found that individuals with an "average" personality most frequently chose an equal distribution. In Table 6 we see, that this result is driven by individuals with an average IQ score. Possibly, individuals with an average intelligence and personality are the most socially adjusted and therefore prone to fulfill social norms like equity in this case.[9] The pattern observed may

[8] 46 out of 87 pupils in this subgroup have an IQ score in the lowest quartile, the remaining 41 in the highest quartile respectively.

[9] Social conformity or adaptability has been shown to be related to personality. See, for example, DeYoung et al. (2002) for one possible transmission mechanism form personality to conformity.

as well be related to risk preferences not surveyed here. Pupils with low/high intelligence and a personality profile that is socially less desirable may be individuals with a high risk preference expressed in high public goods contributions as a risky payoff maximization. Unfortunately, we are not able to investigate this link.

[Include Table 6 here]

So far, we have documented some significant relationships considering endowment variables and other-regarding preferences, indicating at least some scope for the composition hypothesis. Still, number and degree of the correlations are not very high. Also, the direction of correlation is not always suited to support the composition hypothesis leaving room for other possible drivers to explain the difference in experimental decisions by school track.

5.3 School-Track Environment and its Relation to Experimental Outcomes

Subsequently, we investigate whether school-track environment leads to track differences. To investigate this hypothesis, we test for school-track differences in other-regarding preferences for similar pupils across tracks. By similar we refer to the endowment characteristics of IQ and personality traits. We therefore compiled Mann-Whitney-tests for the center 50% of the IQ distribution as well as for the center 40% of the distribution for each of the seven personality traits. Concentrating on the middle of the distributions avoids that outliers drive the results. We find that track-differences are still observable (and significant) if evaluated for these trimmed subgroups. Table 7 shows p-values from Mann-Whitney tests by school track for the subgroups just described. That is, we compare school tracks within the center 50% of the IQ distribution as well as in the center 40% of the distribution of each trait from the JTCI inventory. The general conclusions with respect to tracking from Tables 2 and 3 are still supported: Significant track differences occur in the mean giving for the dictator game as well as in the share of individuals choosing the socially efficient option in the public goods game. The direction of the differences are the same as in Tables 2 and 3. The finding, that track differences also occur for pupils with similar endowment, is a first support for our hypothesis that deviant other-regarding preferences across tracks are due to a treatment effect of the school-track environment.

[Include Table 7 here]

However, it still may be the case that we compare very different individuals from the two tracks. Therefore, conditioning on a larger set of characteristics produces a more robust method to analyze track differences. We refer to the treatment model introduced above and apply propensity score matching to estimate the average treatment effect of the higher school track on other-regarding preferences. Applying this matching procedure is a robust test of the hypothesis that school track influences other-regarding preferences. The average treatment effect (ATE) refers to the effect of a switch from the lower track to the higher track on other-regarding

preferences. The counterfactual needed to calculate the ATE then consists of the hypothetical level of other-regarding preferences if the individual was not enrolled in the higher track but in the lower track instead.

We assume that conditional independence of treatment and $E(y_0)$ holds (ignorability assumption, Rubin, 1978), that is $E(y_0) \perp D|\mathbf{Z}$. In our case this implies that – while controlling for individual endowment – other-regarding preferences are not related to school tracking. As pointed out above, achievement and (mostly unobserved) aspects of the social background drive the tracking decision. we would therefore not expect other-regarding preferences to affect it. We will use the approach of Rosenbaum and Rubin (1983) and condition on the propensity scores instead of the single characteristics to decrease dimensionality. Then, $E(Y) \perp D|P(\mathbf{Z})$. Additionally, overlap is needed, that is $0 < P(D = 1|\mathbf{Z}) < 1$ (see also Rosenbaum and Rubin, 1983). Given these assumptions, the average treatment effect is $E[Y_1 - Y_0|P(\mathbf{Z})]$. To correct for choice-based sampling, we follow the approach of Smith and Todd (2005) and match on the odds-ratio of the propensity score ($P/(1-P)$). Although we only slightly oversampled higher track pupils, results are more reliable when adjusting for choice-based sampling.

The treatment is defined as being in the higher school track. For the propensity score we use IQ scores, personality traits, as well as gender, and having a single parent as determinants of the treatment. Intelligence is the main determinant of the tracking recommendation, which we implement as IQ scores. Besides, personality traits capture aspects of the social behavior, which is also accounted for in the recommendation. Finally, there is evidence, that gender as well as the family background (approximated here by an indicator for living in a single parent household) is crucial for the tracking decision (see Dustmann, 2004). The propensity score is estimated as a probit. Estimates of $P(D|\mathbf{Z})$ are presented in Table 10 in the Appendix. The predictive power of these variables is reflected in a McFadden's-R^2 of .22. Although this is not high, it reflects that the marks that determine the recommendation are only partly related to intelligence as measured by an IQ test.

For the 489 individuals, for whom all variables are available, 398 are within the common support (thereof 189 treated). After matching observations, individuals in the treatment and control group should not differ significantly. We assess the matching quality with the help of the reduction in standardized differences, two-sample t-tests, as well as the McFadden's-R^2 (see Sianesi, 2004, for a description). Table 11 in the Appendix provides the quality tests of the matches. Overall, there is a substantive reduction in the standardized differences. Nevertheless, two-sample t-tests still show some significant differences. The matching quality should thus be evaluated as moderate.

Matching is conducted applying different matching algorithms. Table 8 provides ATE for nearest-neighbor matching (one and three neighbors) with a caliper of .05 (Rosenbaum and Rubin, 1985). Furthermore we report results for radius matching (using the caliper size of .1) and kernel matching using the Epanechnikov kernel. Calculating average treatment effects for the six experimental outcomes already introduced above, we get the following results (see

Table 8). When we control for the selection process of tracking, that is balancing individual characteristics, there still are significant differences in other-regarding preferences with respect to school tracks. Switching from the lower to the higher track would induce pupils to give significantly more in the dictator game (more than .6 tokens more). Regarding the public goods game, there is no significant treatment effect with respect to the mean contribution as well as the share of free-riders. More distinctly, the share of full contributors is significantly lower by about 15%-points for treated individuals. Both effects are robust across matching algorithms.

[Include Table 8 here]

So far, the results show that the observed differences of experimental outcomes by school track are robust to the conditioning on a set of variables. Nevertheless, the matching analysis only conditions on observable characteristics, while there is evidence, that unobservable factors also influence the tracking decision. Tamm (2008), for example, shows that household income is correlated with the tracking decision in Germany, but is not causal to it. Instead, unobserved heterogeneity, like motivation of the parents, is an important influence. In case unobservables play a role, the probability of treatment assignment can differ between two individuals with the same observables but different unobservables. As a sensitivity check for our results, we calculate Rosenbaum Bounds (see Rosenbaum, 2002, chapter 4.2). Therefore, one uses the fact that in case of no unobservable influence on the assignment probability, the odds ratio of the assignment probabilities π of two individuals j and k is equal to one if they have the same observables. We further assume that the odds ratio lies somewhere in between the following interval, with $\Gamma \geq 1$:

$$\frac{1}{\Gamma} \leq \frac{\pi_j(1-\pi_k)}{\pi_k(1-\pi_j)} \leq \Gamma \quad \text{for all } j,k \text{ with } X_j = X_k \tag{5}$$

For $\Gamma = 2$ individuals j and k would differ in their odds of receiving the treatment by factor 2 despite having the same set of observables. Calculating inference with respect to the treatment effect for several values of Γ constitutes a sensitivity check of the possible effect of an unobservable influence. Table 8 provides critical p-values for a test of the hypothesis of no treatment effect. It reveals that both effects are robust to possible unobserved influences. While the effect of the higher school track on dictator giving is robust on an intermediate level (up to a value of $\Gamma = 1.4$), the effect on socially optimal choices in the PGG is very robust. Only if two individuals with the same observables differ in their odds of being in the higher track school by a factor of almost 2, we would assert that there is no treatment effect. Hence, we can conclude, that our result of a significant environmental treatment of school track on other-regarding preferences is largely robust to unobserved heterogeneity.

5.4 Robustness and Validity

5.4.1 Robustness Check for Age-Balanced Sample

It is a common finding that other-regarding preferences of children and adolescents change with age whereby prosocial choices increase with growing age. When comparing pupils from two school-tracks, we therefore have to account for age-differences in the samples. Since the higher track includes pupils from grade 11, aged about 17 years, but the lower track does not, our samples differ with respect to the age-distribution. Differences in other-regarding preferences by school track may therefore also result from differences regarding age. Take for example our result, that higher track students on average give more in the dictator game: with higher giving with increasing age the difference to lower track pupils may simply result from more older pupils in the higher track.

We test for the robustness of our results regarding age by building an age-balanced sample. Randomly ordering students within their track, we restrict the grade-wise subsamples to the lower number of pupils from that grade. Of course, pupils from grade 11 are discarded completely. We thereby end up with the following distribution: 76 pupils in grade 5, 194 in grade 7 and 102 in grade 9. Repeating our analysis for this trimmed sample, we confirm our main results.[10] Dictator giving is higher for pupils in the academic school track, while fully contributing to the public good is more frequently observed for lower track pupils. The latter is highly significant, the former not. Moreover, having an IQ score within the highest quartile of the distribution is associated with lower dictator giving, more frequent egoistic choices and less frequent full contribution in the public goods game. And finally, egalitarian dictator choices are most frequent for pupils with a personality profile that is considered to be moderately socially desirable, while choosing the social optimum in the PGG is most frequently observed for pupils with the least socially desirable personality profile. Both results are significant even in the considerably reduced sample. We therefore conclude, that our main results are not driven by differences in the age-distribution for the school tracks.

5.4.2 Representativeness of the Sample

Since our aim is to explore the variation of other-regarding preferences across subgroups of the population, our sample needs to be representative in order to be externally valid. Table 9 summarizes the comparison of our sample and the respective population values.

[Include Table 9 here]

The overall distribution of IQ scores in the sample is very close to the one in the population, with a mean of 101 and a standard deviation of 15. Also, IQ scores are normally distributed ($p = .18$ for a Shapiro-Wilk test of normality). The track-specific means of IQ scores are 109

[10] Detailed results are available upon request.

for the higher track, and 95 for the lower track. These values are close to the ones provided by Weiß (2006) for pupils from all grades. With respect to the personality traits of the JTCI, factor structure, validity, and correlation of traits in our sample are very close to the results in the representative sample used to scale the inventory. Further characteristics of our sample also point towards representativeness. 19% of the children indicate not living with both parents but in a different household composition. Thereby, the share is 26% in the lower track schools and 11% in the higher track schools. Since a two-parent household includes stepparents in our sample, these figures indicate a distribution close to the actual one, which amounts to a share of 23% of all families being single parent households in 2011 (source: Federal Statistical Office Germany, 2012b). Overall, our aggregate sample is adequately representative of the respective population, especially when compared to student samples, for which most of the analyses of other-regarding preferences are made.

6 Discussion

We refer to tracking as a kind of treatment on other-regarding preferences. Related to this interpretation is an important conclusion: school tracking does not only lead to differences in later outcomes due to a different kind of leaving certificate, but also due to the environmental treatment that a specific school track implies. Besides other-regarding preferences, a multitude of socio-economic outcomes can be differentiated by educational track. Prominent examples are voting behavior (see, e. g., Gallego, 2010), health (see, e. g.,Winkleby et al., 1992), or financial precautions (see, e. g., Lusardi et al., 2010). Educational degree is an indicator of the entity of cognitive and non-cognitive skills that are possibly related to the socio-economic outcome under study. Recent research has shown that the link between school degree and various aspects of labor market success does not solely work through the channel of cognitive skills. Heckman et al. (2006), for example, provide evidence that the causal mechanism from school degree towards labor market success is also transmitted via non-cognitive skills, hence personality traits. In this light, track-specific other-regarding preferences can be interpreted as another aspect of non-cognitive skills.

The relationship of IQ and other-regarding choices is not stable across school-tracks: while higher IQ scores lead to more frequent egoistic choices in the DG for higher track pupils, it relates to less frequent full contributions in the PGG for lower track students. That is, we do not observe a monotonic relationship of intelligence and other-regarding preferences. This result adds to previous findings from Brandstätter and Güth (2002) for adults as well as from Han et al. (2012) for children and adolescents. Brandstätter and Güth (2002) conclude that intelligence is not a valid predictor of other-regarding choices. Han et al. (2012) put forward the hypothesis, that intelligence is related positively with pro-social behavior only for complex task, but not so for simple tasks. Their argumentation is supported by showing that children with IQ scores above 120, between 100 and 120, and below 100 do not choose significantly different allocations in the DG and PGG. Following the distinction between simple and more

complex tasks, we can, however, not rule out, that highly complex other-regarding decision in real life are yet correlated with intelligence. Here, simplification of the behavioral task in the experiment is at the cost of lowering external validity.

An important aspect of our analysis is the direction of the effects. Regarding IQ, we can be sure that school track is not causal for IQ differences. On the one hand, tracking decisions do rely mostly on intelligence, and, on the other hand, intelligence scores are rather stable at the age of 10, which is the youngest age we cover (see Hopkins, 1975). With respect to the personality characteristics measured by the JTCI, direction of the effect is a more difficult issue. We model personality traits as being part of the variables that drive the tracking process. At the same time, we analyze the relationship of personality and other-regarding preferences. If tracking is not solely the result but also a driver of traits, then the correlations presented do not represent a causal result.

One hint for personality traits being a driver not a cause of self-selection could result from looking at the youngest pupils. Fifths graders are part of their track-specific environment only for about 6 month. Hence, we would expect differences in traits to be small if personality is an outcome of tracking. Repeating the analysis from Table 4 for traits in grade 5 only (see Table 12 in the Appendix), we find differences in personality traits of about the same absolute size. Only in case of persistence and cooperativeness are the differences smaller than for the complete sample. Nevertheless, Mann-Whitney tests do not show significance of these differences (not even when they are larger than for the complete sample) due to the smaller sample size and the resulting lower test power. We take this descriptive evidence as an indicator of personality being at least partly a cause of tracking.

Becker et al. (2012) analyze the correlation between economic preference measures – such as risk, time, and other-regarding preferences – and personality traits (Big Five traits and locus of control) for adults. Using experimental as well as survey data, they find only a small number of moderately sized correlations. Discussing these findings, they emphasize that preferences and personality traits are two very different concepts. Their analysis assigns about the same explanatory power to personality traits as well as measures of economic preferences when regressed on representative life outcomes. Hence, Becker et al. (2012) conclude that both sets are complementary constructs. Our results, as presented in the correlation analysis in Table 5 with significant correlations ranging from .09 to .21, reflect their findings.

Relating to previous research on the relationship of personality traits and other-regarding preferences, comparison is impeded by the use of different measures of personality and preferences. Nevertheless, papers by Ben-Ner et al. (2004), Koole et al. (2001), as well as Müller and Schwieren (2012) contain a common finding: the trait of agreeableness relates positively to giving in experiments. De Fruyt et al. (2000) show that the agreeableness trait measured by the Big Five inventory and the cooperativeness character trait from the Temperament and Character Inventory show a significant overlap. Since we were able to show that cooperativeness and giving are positively correlated, our results provide additional support for this link in

a sample of children and adolescents.

Although we use the term school-track environment effect to explain the difference in outcomes, we are aware that the actual causal mechanism may encompass a variety of specific influences that can be subsumed by the notion of environment. The most obvious mechanism is peer influence. This encompasses the transmission mechanism of peers' characteristics or other-regarding preferences influencing the behavioral outcome. Nevertheless direct peer effects with respect to the behavioral outcomes are ruled out by the anonymous design of the experimental procedure. We have also shown that an indirect peer effect via individual endowment does not play an important role. Solely beliefs, for which we do not control, may lead to pupils making different hypotheses about their unknown partners. Beliefs may also vary across school tracks due to a deviant history of experiences in social interactions. Similarly to beliefs, group-specific norms related to ones' social identity may be a driver of track differences. Unfortunately, these are not directly identifiable.

Another possible objection to the existence of a school-track environment effect on other-regarding preferences may come from the literature on intergenerational transmission of human capital. There is growing empirical evidence that preferences, such as risk and trust attitudes (Dohmen et al., 2012), generosity (Wilhelm et al., 2008), as well as ability and non-cognitive skills (Blanden et al., 2007) are transmitted from parents to children. Likewise, contributions to a public good are correlated within the family (Peters et al., 2004). These results are in accordance with micro-economic models depicting population dynamics of traits like altruism (see, e. g., Bisin and Verdier, 2001). In case that other-regarding preferences are linked to socio-economic background, a social segregation of children by school tracking would reproduce intergenerational transmission with respect to the children's preferences. Then, differences in other-regarding preferences by school track would not have a causal interpretation regarding tracking. Nevertheless, there is no evidence on the correlation of other-regarding preferences and socio-economic background. In our data set, living in a single parent household as well as the employment status of the parents are not significantly correlated with other-regarding preferences.[11] We therefore conclude that existing differences in other-regarding preferences by school tracks can be interpreted as a school-track environment effect.

7 Conclusion

Using experimental data, we have shown that children and adolescents vary in their other-regarding preferences. Thereby, differences in altruism, social welfare preferences, and the share of free-riders are distinct across the two school tracks we examine. These deviations cannot be explained by differences in individual endowment, measured by intelligence tests and self-evaluated personality traits.

Despite possible drawbacks, our empirical results point in a very clear direction. Environment,

[11] Not reported. Results are available upon request.

which is represented by school-track, more so than individual endowment, is the important source of variation in other-regarding preferences. Even if personality and school-track environment are simultaneous drivers of other-regarding preferences, stressing school-track environment is a viable conclusion of our analysis. Future research may then be directed at identifying possible channels how a track-specific environment influences outcomes such as other-regarding preferences.

References

Abeler, J. and D. Nosenzo (2013). Self-selection into economics experiments is driven by monetary reward. Discussion Paper 7374, IZA-Institute for the Study of Labor.

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 (2012). 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 15*(2), 1–20.

Austin, E. J., I. J. Deary, M. C. Whiteman, F. G. R. Fowkes, N. L. Pedersen, P. Rabbitt, N. Bent, and L. McInnes (2002). Relationships between ability and personality: Does intelligence contribute positively to personal and social adjustment? *Personality and Individual Differences 32*(8), 1391–1411.

Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012). The relationship between economic preferences and pschological personality measures. Discussion Paper 6470, IZA-Institute for the Study of Labor.

Becker, S. O. and M. Caliendo (2007). mhbounds - Sensitivity analysis for average treatment effects. *Stata Journal 7*(1), 71–83.

Ben-Ner, A., F. Kong, and L. Putterman (2004). Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology 25*(5), 581–589.

Betts, J. R. (2011). The economics of tracking in education. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 3, Chapter 7, pp. 341–382. North-Holland.

Bisin, A. and T. Verdier (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory 97*(2), 298–319.

Blanden, J., P. Gregg, and L. Macmillian (2007). Accounting for intergenerational income persistence: Non-cognitive skills, ability and education. *The Economic Journal 117*(519), C43–C60.

Brandstätter, H. and W. Güth (2002). Personality in dictator and ultimatum games. *Central European Journal of Operations Research 10*(3), 191–215.

Caspi, A. (1998). Personality development across the life course. In W. Damon (Ed.), *Handbook of Child Psychology*, Volume 3, Chapter 6, pp. 311–388. John Wiley & Sons.

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

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.

DeYoung, C. G., J. B. Peterson, and D. M. Higgins (2002). Higher-order factors of the big five predict conformity: Are there neuroses of health? *Personality and Individual Differences 33*(4), 533–552.

DiPrete, T. A. and M. Gangl (2004). Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. Discussion Paper SP I 2004-101, Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2012). The intergenerational transmission of risk and trust attitudes. *The Review of Economic Studies 79*(2), 645–677.

Duncan, S. C., T. E. Duncan, L. A. Strycker, and N. R. Chaumeton (1996). Relations between youth antisocial and prosocial activities. *Journal of Behavioral Medicine 69*(4), 290–307.

Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers 56*(2), 209–230.

Eisenberg, N., R. A. Fabes, and T. L. Spinrad (2006). Prosocial development. In N. Eisenberg (Ed.), *Handbook of Child Psychology* (6th ed.), Volume 3, Chapter 11, pp. 646–718. John Wiley & Sons.

Federal Statistical Office Germany (2012a). Allgemeinbildende Schulen, Fachserie 11, Reihe 1.

Federal Statistical Office Germany (2012b). Bevölkerung und Erwerbstätigkeit, Fachserie 1, Reihe 3.

Federal Statistical Office Germany (2012c). Regionalstatistik, Beschäftigte.

Federal Statistical Office Germany (2012d). Regionalstatistik, Bevölkerung.

Fehr, E., H. Bernhard, and B. Rockenbach (2008). Egalitarianism in young children. *Nature 454*, 1079–1083.

Gallego, A. (2010). Understanding unequal turnout: Education and voting in comparative perspective. *Electoral Studies 29*(2), 239–248.

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

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

Han, R., J. Shi, W. Yong, and W. Wang (2012). Intelligence and prosocial behavior: Do smart children really act nice? *Current Psychology 31*(1), 88–101.

Harbaugh, W. T. and K. Krause (2000). Children's altruism in public good and dictator experiments. *Economic Inquiry 38*(1), 95–109.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Hindriks, J., M. Verschelde, G. Rayp, and K. Schoors (2010). Ability tracking, social segregation and educational opportunity: Evidence from Belgium. Discussion Paper 2010/82, Center for Operations Research and Econometrics (CORE).

Hopkins, K. D. (1975). Ten-year stability of verbal and nonverbal IQ scores. *American Educational Research Journal 12*(4), 469–477.

John, K. and S. L. Thomsen (2013). Environment and other-regarding preferences. Discussion Paper 4, Niedersächsisches Institut für Wirtschaftsforschung.

Koole, S. L., W. Jager, A. E. van den Berg, C. A. J. Vlek, and W. K. B. Hofstee (2001). On the social nature of personality: Effects of extraversion, agreeableness, and feedback about collective resources use on cooperation in a resource dilemma. *Personality and Social Psychology Bulletin 27*(3), 289–301.

Kremer, M. and D. Levy (2008). Peer effects and alcohol use among college students. *Journal of Economic Perspectives 22*(3), 189–206.

Leuven, E. and B. Sianesi (2012). psmatch2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *Statistical Software Components*.

Lusardi, A., O. S. Mitchell, and V. Curto (2010). Financial literacy among the young: Evidence and implications for consumer policy. *Journal of Consumer Affairs 44*(2), 358–380.

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.

Meeus, W., A. Oosterwegel, and A. Vollebergh (2002). Parental and peer attachment and identity development in adolescence. *Journal of Adolescence 25*(1), 93–106.

Mühlenweg, A. M. and P. A. Puhani (2010). The evolution of the school-entry age effect in a school tracking system. *The Journal of Human Resources 45*(2), 407–438.

Müller, J. and C. Schwieren (2012). What can the Big Five personality factors contribute to explain small-scale economic behavior. Discussion Paper 2012-028/1, Tinbergen Institute.

Murninghan, J. K. and M. S. Saxon (1998). Ultimatum bargaining by children and adults. *Journal of Economic Psychology 19*, 415–445.

OECD (2012). Equity and equality in education: Supporting disadvantaged students and schools. Report, OECD.

Peters, H. E., A. S. Ünür, J. Clark, and W. D. Schulze (2004). Free-riding and the provision of public goods in the family: A laboratory experiment. *International Economic Review 45*(1), 283–299.

Rosenbaum, P. R. (2002). *Observational Studies*. Springer.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Rosenbaum, P. R. and D. B. Rubin (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician 39*(1), 33–38.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5), 688–701.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics 6*(1), 34–58.

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.

Sianesi, B. (2004). An evaluation of the swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics 86*(1), 133–155.

Smith, J. A. and P. E. Todd (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics 125*(1-2), 305–353.

Tamm, M. (2008). Does money buy higher schooling? Evidence from secondary school track choice in Germany. *Economics of Education Review 27*(5), 536–545.

Vesterlund, L. (2006). Why do people give? In W. W. Powell and R. Steinberg (Eds.), *The Nonprofit Sector: A Research Handbook*, Chapter 24, pp. 568–587. Yale University Press.

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 (CFT 20R)*. Hogrefe.

Wilhelm, M. O., E. Brown, P. M. Rooney, and R. Steinberg (2008). The intergenerational transmission of generosity. *Journal of Public Economics 92*(10), 2146–2156.

Winkleby, M. A., D. E. Jatulis, E. Frank, and S. P. Fortmann (1992). Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *American Journal of Public Health 82*(6), 816–820.

A Appendix

Table 1: Description of Personality Dimensions in the JTCI

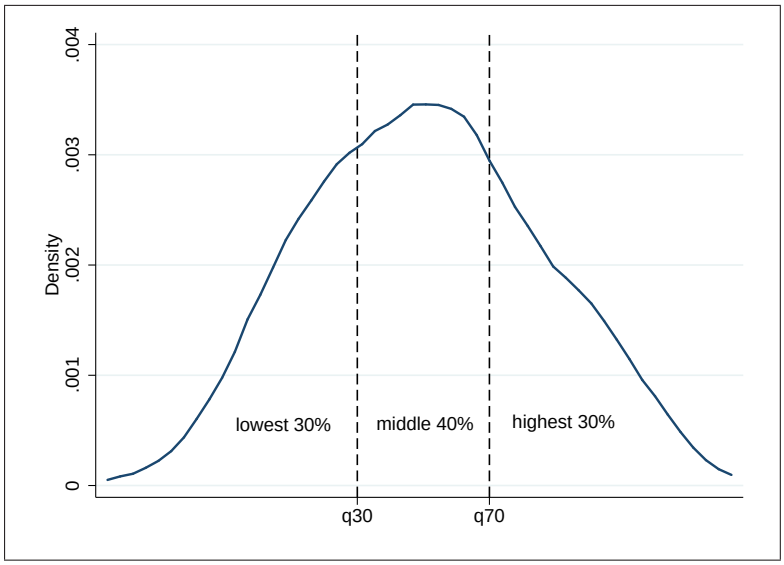
Trait	Concept	Adjectives
TEMPERAMENT		
Novelty Seeking	relates to the behavioral activation system; extent of exploratory activity, approach to novel stimuli, and avoidance of averse stimuli	impulsive, exploratory, curious versus reflective, orderly
Harm Avoidance	relates to the behavioral inhibition system; extent of being shy, fatigable, fearful, or pessimistic	shy, worrying, pessimistic versus confident, outgoing, optimistic
Reward Dependence	relates to the behavioral dependence system; extent of regulating one's behavior due to social cues	attached, sentimental, warm versus detached, withdrawn, cold
Persistence	relates to the behavioral persistence system; the intrinsic maintenance of a behavior	hard-working, ambitious versus gives up easily, inactive
CHARACTER		
Self-Directedness	self-concept; relates to responsibility and resourcefulness in order to achieve personal goals	responsible, reliable versus immature, unreliable
Cooperativeness	extent of feeling as part of the social system; comprises social tolerance, empathy, helpfulness, compassion and moral principles	empathic, tolerant, helpful versus revengeful, intolerant
Self-Transcendence	extent of feeling as part of one's surroundings as a unitive whole; tendency of creative inspiration and spirituality	patient, creative, self-forgetful versus impatient, self-conscious

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.

Figure 1: Location of Federal States where Pupils are surveyed

Data has been surveyed in the two shaded federal states, the western being Lower Saxony, the eastern being Saxony-Anhalt.

Figure 2: Desirability of Personality



Kernel density estimate of the sum of deviations from the median traits (Epanechnikov-kernel, bandwidth=25.0). Partition into three types of personality is shown: The lowest 30%, the middle 40%, and the highest 30%.

Table 2: Summary of Results for Dictator Game

	N	**Contribution (of 10)**	**Contribution**= 0	**Equal Split**
Overall	530	2.23	.19	.08
		(1.90)	(.40)	(.26)
Lower Track	294	2.00	.21	.06
		(1.79)	(.41)	(.24)
Higher Track	236	**2.54**	.17	.09
		(1.98)	(.38)	(.29)
IQ (lowest quartile)	144	2.26	.19	.06
		(1.97)	(.40)	(.23)
IQ (second quartile)	148	2.20	.17	.08
		(1.89)	(.38)	(.27)
IQ (third quartile)	107	2.39	.15	.09
		(1.78)	(.36)	(.29)
IQ (highest quartile)	125	2.08	**.26**	.07
		(1.92)	(.44)	(.26)
Personality score low	152	**1.88**	*.24*	**.03**
		(1.78)	(.43)	(.16)
Personality score medium	209	*2.42*	*.16*	**.11**
		(1.96)	(.37)	(.31)
Personality score high	151	2.30	.19	.05
		(1.87)	(.39)	(.23)

Displayed are means, standard errors in brackets. Significant differences across groups from two-sample Mann-Whitney tests at the 5%-significance level are displayed in bold face, at the 10%-significance level in italics. Differences between IQ quartiles and personality dimensions have been calculated against the respective rest of the sample, that is e. g. first quartile against quartiles 2, 3, and 4.

Table 3: Summary of Results for Public Good Game

	N	**Contribution (of 10)**	**Contribution**= 0	**Contribution**= 10
Overall	530	4.39	.08	.12
		(2.93)	(.28)	(.32)
Lower Track	294	4.46	.11	.17
		(3.22)	(.31)	(.37)
Higher Track	236	4.31	**.06**	**.05**
		(2.52)	(.23)	(.22)
IQ (lowest quartile)	144	*4.83*	.08	**.17**
		(3.13)	(.27)	(.38)
IQ (second quartile)	148	4.37	.11	.12
		(2.96)	(.32)	(.32)
IQ (third quartile)	107	3.97	.08	*.07*
		(2.66)	(.26)	(.25)
IQ (highest quartile)	125	4.38	.06	.10
		(2.79)	(.25)	(.30)
Personality score low	152	4.76	**.05**	**.17**
		(3.16)	(.21)	(.37)
Personality score medium	209	4.43	.09	.11
		(2.88)	(.29)	(.31)
Personality score high	151	4.13	.08	*.07*
		(2.65)	(.27)	(.26)

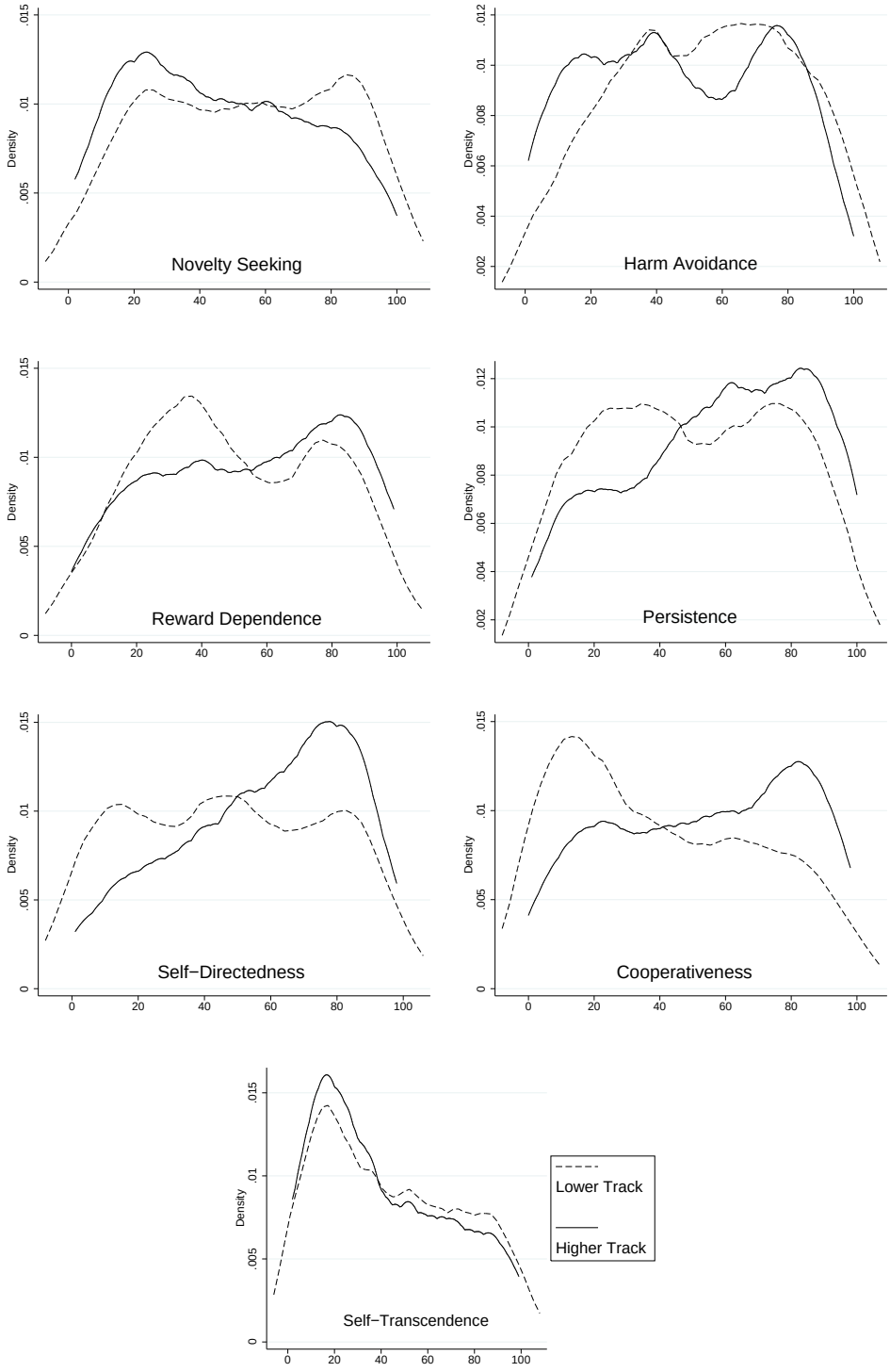
Displayed are means, standard errors in brackets. Significant differences across groups from two-sample Mann-Whitney tests at the 5%-significance level are displayed in bold face, at the 10%-significance level in italics. Differences between IQ quartiles and personality dimensions have been calculated against the respective rest of the sample, that is e. g. first quartile against quartiles 2, 3, and 4.

Table 4: Means of individual characteristics by school track

Variable	Lower Track	Higher Track	(1)	(2)
IQ	95.3	108.7	.00	.00
Novelty Seeking	53.8	46.7	.01	.03
Harm Avoidance	54.1	48.1	.02	.12
Reward Dependence	50.1	55.0	.05	.10
Persistence	50.0	57.1	.00	.02
Self-Directedness	47.1	57.2	.00	.00
Cooperativeness	40.5	53.3	.00	.00
Self-Transcendence	44.2	41.1	.23	.46
Share of girls	.53	.53	.84	.99
Single parent	.22	.08	.00	.01

Column (1) displays p-values of a two-sample Mann-Whitney-tests on equal means. Column (2) displays p-values of a Kolmogorov-Smirnov test on equal distributions. Number of observations varies by variables: N=524 for IQ, N=512 for personality traits.

Figure 3: Differences in Distributions of Personality Characteristics by School Track



Kernel density estimates using the Epanechnikov-Kernel and bandwidth=8.0

Table 5: Spearman's Correlation Coefficients for Endowment and Other-Regarding Preferences

		DG	**DG=0**	**DG=5**	**PGG**	**PGG=0**	**PGG=10**
IQ	Overall						-.12
	Lower Track				-.15		
	Higher Track		.15				
Novelty Seeking	Overall	-.15					
	Lower Track			-.14			
	Higher Track	-.13					
Harm Avoidance	Overall				.09		
	Lower Track						
	Higher Track						
Reward Dependence	Overall		-.11				-.11
	Lower Track						
	Higher Track		-.14				-.14
Persistence	Overall				-.15	.13	-.11
	Lower Track				-.13		
	Higher Track				-.19	.21	-.19
Self-Directedness	Overall						
	Lower Track						
	Higher Track						
Cooperativeness	Overall	.15		.10			-.11
	Lower Track						
	Higher Track	.15	-.14				
Self-Transcendence	Overall						
	Lower Track						
	Higher Track						-.14

Displayed are Spearman's rank correlation coefficients significant at the 5%-significance level.

Table 6: Summary of Results for all Games by IQ and Personality

		N	**DG**	**DG**= 0	**DG**= 5	**PGG**	**PGG**= 0	**PGG**= 10
Overall		530	2.23	.19	.08	4.39	.08	.12
			(1.90)	(.40)	(.26)	(2.93)	(.28)	(.32)
IQ 25-75%	(1)	65	2.14	.20	.03	4.23	.08	.15
			(2.07)	(.40)	(.17)	(3.25)	(.27)	(.36)
	(2)	105	2.44	.11	*.13*	4.29	.11	.08
			(1.77)	(.32)	(.34)	(2.77)	(.32)	(.27)
	(3)	78	2.14	.19	.04	4.17	.08	.08
			(1.73)	(.40)	(.19)	(2.62)	(.27)	(.27)
IQ 1st&4th quart.	(1)	87	*1.69*	.28	.02	*5.16*	*.02*	.17
			(1.50)	(.45)	(.15)	(3.04)	(.15)	(.38)
	(2)	104	2.39	.20	.09	4.57	.07	.14
			(2.15)	(.40)	(.28)	(2.99)	(.25)	(.35)
	(3)	73	2.47	.18	.07	4.10	.08	.07
			(2.00)	(.39)	(.25)	(2.70)	(.28)	(.25)

Displayed are means, standard errors in brackets. (1), (2), (3) refer to the lower 30%, the middle 40%, and the upper 30% for the distribution of the sum of deviations from the median traits of the JTCI. Lower values (1) (higher values (3)) thereby refer to a less (more) socially desirable personality type. Significant differences across groups from two-sample Mann-Whitney tests at the 5%-significance level are displayed in boldface. Differences between IQ quartiles and personality dimensions have been calculated against the respective rest of the sample, that is e. g. first quartile against quartiles 2, 3, and 4.

Table 7: p-Values for Track Differences in Center Populations

	N	**DG**	**DG**= 0	**DG**= 5	**PGG**	**PGG**= 0	**PGG**= 10
IQ	269	.00	.08	.14	.07	.07	.08
Novelty Seeking	214	.01	.74	.41	.58	.21	.11
Harm Avoidance	234	.00	.08	.05	.67	.12	.00
Reward Dependence	216	.26	.89	.53	.80	.60	.00
Persistence	228	.00	.02	.45	.30	.33	.01
Self-Directedness	201	.00	.15	.15	.47	.08	.01
Cooperativeness	195	.03	.53	.44	.88	.32	.12
Self-Transcendence	177	.06	.77	.76	.30	.26	.01

Displayed are p-values from two-sample Mann-Whitney tests by school track. The subgroups refer to center parts of the distributions: for IQ, pupils within the second and third quartile of the distribution are included. For the personality traits, the center 40% of the distribution is included.

Table 8: Average Treatment Effects for Higher School Track

	DG	**DG=0**	**DG=5**	**PGG**	**PGG=0**	**PGG=10**
NN(1), caliper=.05	.64***	-.08*	.01	-.29	-.05	-.15***
	(.21)	(.05)	(.03)	(.40)	(.03)	(.04)
$p(\Gamma = 1.0)$	.00	.27	.28	.07	.12	.00
$p(\Gamma = 1.1)$	.00	.36	.34	.18	.15	.01
$p(\Gamma = 1.2)$	.01	.45	.41	.35	.19	.01
$p(\Gamma = 1.3)$	.04	.54	.47	.53	.22	.02
$p(\Gamma = 1.4)$	.09	.54	.53	.70	.26	.02
$p(\Gamma = 1.5)$	.17	.46	.58	.82	.30	.03
$p(\Gamma = 1.6)$	.28	.40	.57	.90	.33	.04
$p(\Gamma = 1.7)$	.41	.34	.53	.95	.36	.06
$p(\Gamma = 1.8)$	.54	.28	.48	.98	.40	.07
$p(\Gamma = 1.9)$	.65	.24	.44	.99	.43	.08
$p(\Gamma = 2.0)$	.75	.20	.40	.99	.46	.10
NN(3), caliper=.05	.74***	-.07	.03	-.17	-.06	-.14***
	(.21)	(.05)	(.03)	(.39)	(.03)	(.04)
Radius, caliper=.1	.79***	-.06	.03	-.12	-.05	-.14***
	(.16)	(.04)	(.02)	(.35)	(.03)	(.04)
Kernel (Epanechnikov)	.66***	-.06	.01	-.40	-.04	-.17***
	(.19)	(.05)	(.03)	(.41)	(.03)	(.05)

Displayed are average treatment effects (ATE) after propensity score matching with higher school track as the treatment. Results have been obtained using STATA procedure *psmatch2* by Leuven and Sianesi (2012). Standard errors in parentheses are bootstrapped with 200 replications. *** denotes significance at the 1%-level, ** at the 5%-level, and * at the 10%-level. Matching is on the odds ratio of the propensity score. P-values for Rosenbaum Bounds have been obtained using *rbounds* (DiPrete and Gangl, 2004) for continuous outcomes and *mhbounds* (Becker and Caliendo, 2007) for binary outcomes.

Table 9: Representativeness of the Sample

Variable	Sample Mean	Population Value
Share of girls[a]	53%	49%
Share in higher track[b]	46%	39%
IQ[c]	101.3	100.0
IQ (higher track)[c]	108.7	109.0
IQ (lower track)[c]	95.3	103.0
Novelty Seeking[d]	50.6	50.0
Harm Avoidance[d]	51.3	50.0
Reward Dependence[d]	52.3	50.0
Persistence[d]	53.2	50.0
Self-Directedness[d]	51.7	50.0
Cooperativeness[d]	46.3	50.0
Self-Transcendence[d]	42.8	50.0
Age[d]	13.2	13.5
Share not living with both parents[a]	19%	23%

[a] Federal Statistical Office Germany (2012b)
[b] Federal Statistical Office Germany (2012a)
[c] Validity sample of CFT, Weiß (2006)
[d] Validity sample of JTCI, Goth and Schmeck (2009)

Table 10: Probit Estimates of the Propensity Score

Variable	Coeff.	s.e.
IQ	0.048***	0.005
Novelty Seeking	-0.000	0.003
Harm Avoidance	-0.003	0.003
Reward Dependence	0.000	0.003
Persistence	-0.005*	0.003
Self-Directedness	0.003	0.003
Cooperativeness	0.012***	0.003
Self-Transcendence	0.000	0.002
Female	-0.221	0.143
Single parent	-0.474**	0.189
Constant	-5.106***	0.633

Coefficients and standard errors from a probit regression are presented. *** denotes significance at the 1%-level, ** at the 5%-level and * at the 10%-level. N=489, McFadden's-R^2=0.22.

Table 11: Matching Quality

		Mean		Reduction in stand. Differences		t-test	
Variable	Sample	Treated	Control	%stand. Difference	|Difference|	t	$p > \|t\|$
IQ	Unmatched	108.6	95.4	101.2		11.18	0.00
	Matched	104.8	104.2	4.0	96.1	-5.01	0.00
Novelty Seeking	Unmatched	47.1	53.6	-23.2		-2.56	0.01
	Matched	49.2	42.3	24.2	-4.2	3.43	0.00
Harm Avoidance	Unmatched	48.1	54.6	-23.4		-2.59	0.01
	Matched	50.6	50.4	0.7	96.9	0.77	0.44
Reward Dependence	Unmatched	55.1	50.9	14.7		1.63	0.10
	Matched	54.7	55.1	-1.7	88.4	-1.23	0.22
Persistence	Unmatched	56.8	50.6	21.9		2.41	0.02
	Matched	55.3	58.4	-10.9	50.2	-2.44	0.02
Self-Directedness	Unmatched	57.2	47.4	35.2		3.87	0.00
	Matched	55.4	58.4	-10.8	69.3	-3.79	0.00
Cooperativeness	Unmatched	53.1	40.9	42.2		4.65	0.00
	Matched	50.5	52.5	-6.9	83.6	-2.96	0.00
Self-Transcendence	Unmatched	41.1	44.5	-11.5		-1.26	0.21
	Matched	41.7	39.5	7.5	34.4	2.05	0.04
Female	Unmatched	.53	.55	-3.6		-0.39	0.70
	Matched	.51	.50	1.1	70.3	-0.20	0.85
Single parent	Unmatched	.08	.22	-38.1		-4.16	0.00
	Matched	.10	.09	3.0	92.1	1.66	0.10

Output generated using Stata program *pstest* after *psmatch2* by Leuven and Sianesi (2012). %stand. Difference refers to the standardized difference of covariates across the two groups, and is equal to $100 \times (\bar{X}_1 - \bar{X}_0)/\sqrt{0.5 \times (V_1(X) + V_0(X))}$ with V as the variances in the respective track.

Table 12: Means of Individual Characteristics by School Track in Grade 5

Variable	Lower Track	Higher Track	(1)	(2)
IQ	95.0	111.7	.00	.00
Novelty Seeking	47.3	40.4	.21	.41
Harm Avoidance	55.0	46.1	.14	.19
Reward Dependence	55.8	60.6	.23	.29
Persistence	64.5	69.1	.13	.12
Self-Directedness	43.6	60.0	.01	.06
Cooperativeness	48.5	57.2	.17	.21
Self-Transcendence	47.4	38.5	.14	.33

Column (1) displays p-values of a two-sample Mann-Whitney-tests on equal means. Column (2) displays p-values of a Kolmogorov-Smirnov test on equal distributions. Number of observations varies by variables: N=118 for IQ, N=100 for personality traits.