The Formation of Prosociality: Causal Evidence on the Role of Social Environment

Fabian Kosse (LMU Munich)
Thomas Deckers (University of Bonn)
Pia Pinger (University of Bonn)
Hannah Schildberg-Hörisch (DICE)
Armin Falk (University of Bonn)

Discussion Paper No. 167

July 30, 2019

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

The Formation of Prosociality: Causal Evidence on the Role of Social Environment*

Fabian Kosse[1,2], Thomas Deckers[3], Pia Pinger[2,3], Hannah Schildberg-Hörisch[4] and Armin Falk[2,3]

[1] Ludwig-Maximilians-University Munich (LMU)
[2] Institute on Behavior and Inequality (briq)
[3] University of Bonn
[4] University of Düsseldorf

Abstract

This study presents descriptive and causal evidence on the role of social environment for the formation of prosociality. In a first step, we show that socioeconomic status (SES) as well as the intensity of mother-child interaction and mothers' prosocial attitudes are systematically related to elementary school children's prosociality. In a second step, we present evidence on a randomly-assigned variation of the social environment, providing children with a mentor for the duration of one year. Our data include a two-year follow-up and reveal a significant and persistent increase in prosociality in the treatment relative to the control group. Moreover, enriching the social environment bears the potential to close the observed gap in prosociality between low and high SES children. A mediation analysis of the observed treatment effect suggests that prosociality develops in response to stimuli in the form of prosocial role models and intense social interactions.

Keywords: Formation of preferences, prosociality, social preferences, trust, social inequality.

JEL-Codes: D64, C90

*Financial support through the Leibniz Programme of the German Science Foundation (DFG), the Benckiser Stiftung Zukunft, and the German Research Foundation through CRC TR 190 and CRC TR 224 is gratefully acknowledged.

1 Introduction

Prosociality is a particularly important aspect of human personality and it affects a wide range of economic decisions and outcomes, e.g., the provision of public goods, contract enforcement, management of commons, governmental and judicial efficiency or economic growth (Knack and Keefer, 1997; La Porta et al., 1997; Fehr and Gächter, 2002; Henrich et al., 2004; Ostrom et al., 2002; Guiso, Sapienza, and Zingales, 2009; Cooper and Kagel, 2009; Burks et al., 2016). Prosocial behavior is not only a crucial factor for the functioning of societies but also an important skill that affects health, well-being, and labor market success (Dohmen et al., 2009; Carpenter and Seki, 2011; Becker et al., 2012; Algan et al., 2016; Deming, 2017).[1]

The importance of prosociality in economic contexts is thus in line with findings from a vast body of empirical literature on the returns to non-cognitive skills (see, e.g., Heckman, Stixrud, and Urzua, 2006). It is also reflected in economic theory, which, pioneered by Becker's work on altruism (Becker, 1976), incorporates other-regarding preferences and beliefs as key components (e.g., Rabin, 1993; Fehr and Schmidt, 1999; Charness and Rabin, 2002; Bénabou and Tirole, 2006; Falk and Fischbacher, 2006; Fehr and Schmidt, 2006).

Despite its fundamental importance, and significant advances in understanding the *consequences* of prosociality, little is known about how prosociality *forms*. This paper therefore provides evidence on the formation of prosociality in children with a particular focus on the role of the social environment. It documents *that* prosociality in children is malleable and *how* it can be enhanced.

We identify *prosocial attachment figures* and *intense social interactions* as two main drivers of child prosociality.[2] Prosocial attachment figures are highly prosocial individuals with a close social connection to the child. They enable children to observe and imitate a prosocial role model and thereby internalize prosocial beliefs, preferences and behaviors (Skinner, 1953; Williamson, Donohue, and Tully, 2013). Intense social interactions are joint activities that foster social reinforcement and feedback (Eisenberg, Fabes, and Spinrad, 2006; Eisenberg, Spinrad, and Knafo-Noam, 2015). In order to obtain exogenous variation in these stimuli, we randomly assigned a sample of elementary school children to an enriched social environment in form of a mentoring program. This mentoring program is a well-established non-

[1] Using data from the German Socio-Economic Panel (SOEP), we find for example that a one standard deviation increase in prosociality translates into 0.45 standard deviations improved subjective well-being, a two percentage point reduction in the probability of being unemployed, as well as higher wages, and better health (see Becker et al. (2012) and table B1).

[2] For biological and genetic determinants of prosociality, see e.g. Fehr and Fischbacher (2003), Kosfeld et al. (2005) and Cesarini et al. (2008, 2009).

profit program called "Balu und Du" (German for "Baloo and You"). It provides children with a volunteer mentor for the duration of one year. Conceptually, the idea of the program is to extend a child's horizon and foster the acquisition of new skills and experiences through intense social interactions between mentor and child. During the intervention, the child experiences an unrelated and highly prosocial attachment figure taking responsibility and devoting effort and time with him/her.

Before and after the intervention, children and their mothers were interviewed by trained interviewers. Children participated in incentivized choice experiments and answered a short questionnaire. Mothers completed an extensive questionnaire covering socio-economic background information, interaction patterns and assessments of personality regarding their children and themselves. In order to yield a comprehensive measure of a child's prosociality, we collected data on three main facets: altruism, trust and other-regarding behavior in everyday life. Altruism reflects prosocial motivation, trust indicates prosocial beliefs, and other-regarding behavior in everyday life appraises prosocial behaviors. The corresponding facets have been shown to predict real-life societal and individual success (Becker et al., 2012; Falk et al., 2018). The respective measures combine information from incentivized experiments and statements from questionnaires. Altruism summarizes children's choices in three incentivized dictator games. Trust is measured using a well-established, age-adapted and experimentally-validated three-item trust questionnaire administered to the child. Moreover, we asked mothers to assess their child's other-regarding behavior in everyday life using the prosociality subscale of the widely used Strength and Difficulties Questionnaire (SDQ) (Goodman, 1997). Our measures thus provide a broad-ranging characterization of prosocial disposition using different data collection methods and statements from children and mothers. This approach reduces measurement error and potential demand effects (Hertwig and Ortmann, 2001). For the main analysis, we collapse all three facets into one joint measure of children's prosociality, but we also provide results for each facet separately. In addition, to understand the respective role of attachment figures, we elicit prosociality for mothers and mentors. As for children, this measure consists of the three facets altruism, trust and other-regarding behavior.

Our sample was recruited using official registry data. It comprises families interested in participating in the mentoring intervention and the interviews. We study three distinct groups that differ in terms of socio-economic and treatment status. Based on socio-demographic background information (household income, parental education and single-parent status), we classify families as either low or high SES. Among low SES families, we randomly assigned a subset of families to participate in

the intervention, reflecting the treatment group (Treatment Low SES). The remaining families with a low SES background form our intervention control group (Control Low SES). The third group comprises families with high SES background (Control High SES). Differences in social environment and child prosociality among Control Low SES and Control High SES families allow us to derive and test hypotheses regarding the two potential drivers of prosociality formation, prosocial attachment figures and intense social interaction, respectively. Comparing Control Low SES and Treatment Low SES provides causal evidence on how providing children with an enriched social environment can promote their prosociality. Overall, we collected three waves of data. After a baseline wave of interviews (wave 1), the treatment was randomly assigned and implemented. The one-year treatment period was closely followed by a post-treatment wave of interviews (wave 2) and a two-year follow-up wave (wave 3). In the main analysis, we focus on post-treatment wave 2 data, but also use wave 1 to study baseline balance and systematic attrition, as well as wave 3 data to investigate whether the observed effects are enduring.

Our main findings can be summarized as follows. Using the Control High SES and Control Low SES groups, we first document a pronounced SES gap in prosociality among elementary school children, amounting to 22.6% of a standard deviation. Moreover, we show that low and high SES families differ in their social environment and provide suggestive evidence that alterable stimuli in form of prosocial attachment figures and intense social interaction explain a substantial part of the observed SES gap in prosociality. This sets the stage for our main analysis: our first key result demonstrates that prosociality in elementary school children is indeed malleable and can be enhanced. Children who were offered participation in a mentoring intervention to enrich their social environment score 27.3% of a standard deviation higher on the prosociality measure than children from the control group. This effect is sizable regarding the program's intensity (about 92 hours on average per year) and costs (about 1,000 EUR per child per year).[3] Moreover, the observed high-low SES developmental gap in prosociality is closed, i.e., children from Treatment Low SES and Control High SES score very similarly on the prosociality measure. Using wave 3 data, we can show that the effect of the mentoring program is enduring. Two years after the end of the intervention, we observe a general increase in prosociality for all three groups,[4] and, most importantly, persistence of the high-low SES gap

[3]Using these numbers and further analyses we estimate an internal rate of return of around 8.5%, see section 4 for details.

[4]This finding is in line with prior cross-sectional evidence on developmental patterns of prosociality, see e.g., Harbaugh and Krause (2000), Sutter and Kocher (2007), Almås et al. (2010), Fehr, Bernhard, and Rockenbach (2008); Fehr, Glätzle-Rützler, and Sutter (2013).

and the treatment effect. Prosociality in Treatment Low SES is thus significantly higher than in Control Low SES, and there is no significant difference between Treatment Low SES and Control High SES. A particularly important question concerns the underlying mechanisms of the mentoring intervention. In a mediation analysis, we show that roughly 40% of the treatment effect is due to exposure to a highly prosocial attachment figure and another 20% is due to an increase in intense social interactions. Linking the process of prosociality formation to the program's main components thus provides more general evidence on how prosociality in elementary school children forms and how it can be enhanced.

The contribution of this paper is threefold. First, our findings demonstrate the malleability of individual prosociality by showing that a moderate enhancement of the social environment can have substantive and enduring effects. Our results thus constitute a proof-of-concept that prosocial motivations, beliefs and actions can be changed, and that elementary school age is a sensitive period in this respect. Second, we document that prosociality forms in response to a child's attachment figures and intense social interactions.[5] Exogenous, intervention-induced variation in these stimuli allows uncovering and discussing their overall and relative importance for the formation of child prosociality. Our findings thus relate to literature on skill formation (see e.g. Cunha and Heckman, 2007) that highlights childhood as a critical and sensitive period and identifies particular investments as primary drivers of personality formation. Third, by evaluating the mentoring program that formed the basis of the randomized control trial (RCT), we add to the literature on interventions for elementary school children (Rodríguez-Planas, 2012; Kautz et al., 2014). Thus far, large-scale and long-term evaluations of interventions at elementary school age that are based on RCTs are rare, mostly focusing on in-school programs, and their evidence is mixed (for an overview and discussion, see Kautz et al., 2014).[6] While some programs show positive effects on health and functioning in school (Hawkins et al., 1999, 2005, for example) or patience (Alan and Ertac, 2018), others display adverse long-run effects (McCord, 1978). Instead, our results indicate that informal out-of-school mentoring during elementary school age has the potential to systematically affect character formation and close developmental gaps arising from "the

[5]The importance of the social environment for child development at young ages has been documented, for example, in Nelson et al. (2007), Heckman et al. (2010b), Attanasio et al. (2014), Campbell et al. (2014) and Doyle et al. (2017).

[6]Wider literature exists on mentoring and in particular the Big Brothers Big Sisters program (Foster, 2001; Eby et al., 2008; Moodie and Fisher, 2009). However, few of these studies comprise randomization and long-run follow ups. Moreover, this literature almost exclusively focuses on adolescent youths (see, e.g. Grossman and Tierney, 1998). For studies in the school-context focusing on prosociality see Rao (2014); John and Thomsen (2015); Cappelen et al. (2016); Alan and Ertac (2017).

accident of birth" (Heckman, 2008, 2013). In this sense, our findings also inform the discussion on social mobility, societal inequality and the intergenerational persistence of life outcomes (Case, Lubotsky, and Paxson, 2002; Currie and Moretti, 2003; Aizer and Currie, 2014; Putnam, 2015).

The remainder of the paper is organized as follows. In the next section we discuss recruitment, our measures, details of the mentoring intervention and its implementation. Section 3 first presents descriptive evidence and deduces more specific research questions regarding the relationship between a child's social environment and formation of prosociality. Subsequently, it unveils the causal effect of enriching a child's social environment on prosociality and delineates the underlying mechanisms. Finally, Section 4 concludes.

2 Study Design

This section introduces the design of the study. We first report how children and their families were recruited, how we classified them in terms of SES and how children were randomly assigned to treatment, resulting in our three groups Control Low SES, Treatment Low SES and Control High SES, respectively. We then describe the setting and details of the interviews, including a description of our survey and experimental measures. Finally, we discuss the scope, concept and procedures of the mentoring intervention.

2.1 Recruitment of sample

Figure 1 presents a flow chart of the timing, sampling and procedural details of our study. Recruiting started in summer 2011. We used official registry data to obtain more than 95% of the addresses of families (with children aged from seven to nine) living in the German cities of Bonn and Cologne. Invitations to take part in the mentoring program and the interviews were sent to all families with children born between 09/2003 and 08/2004 and to one-third of the families with children born between 09/2002 and 08/2003. Birth dates were chosen such that children of the younger cohort were typically in second grade.[7] In summer 2011, families were contacted via postal mail. We announced the possibility to take part in the mentoring program and the interviews. We informed parents that participation in the mentoring program was not guaranteed due to limited capacity (which was the

[7]Nearly all children in our sample (99%) attend public schools. The quality of public elementary schools in Germany is very homogeneous. Teachers are typically centrally allocated to schools, have obtained the same qualifications and are paid the same salaries.

case). We asked them to send back a short questionnaire concerning socio-economic characteristics of the household, and to sign a non-binding letter of intent to take part in the interviews and the mentoring program, if interested. We received 1,626 complete and valid responses.[8] Using responses to the questionnaire, we categorized households as either low or high SES households, respectively. SES reflects the level of resources available at the household level, i.e., material, educational and time resources. Accordingly, a household was classified as low SES if at least one of the three following criteria was met: (i) *Low income*: Equivalence income of the household is lower than 1,065 Euro. This corresponds to the 30% quantile of the German income distribution.[9] (ii) *Low education*: Neither the mother nor the father of the child have a school-leaving degree qualifying for university studies. (iii) *Single-parent status*: A parent is classified as a single parent if he/she is not living together with a partner.[10] Households for which none of the three criteria applied were classified as high SES.

Low SES households form our target group and therefore we invited all low SES families to take part in the study. In order to be eligible for treatment, they had to participate in the first wave of interviews (fall 2011) and provide written consent to allow the transmission of their addresses to the organization running the mentoring program. Out of 590 eligible low SES families, 212 were randomly selected and constitute our intention-to-treat (ITT) group (Treatment Low SES). The remaining 378 families form the control group (Control Low SES). We used stratified random treatment assignment: stratification considered fourteen sub-groups resulting from the combination of city (Bonn or Cologne) and SES criteria (low income and/or low education and/or single-parent status). Stratification was used to ensure a proportional representation of all combinations of criteria in the ITT group, and that the number of selected children matched the local supply of mentors.[11] After treatment assignment, we transmitted contact information of the ITT group to the mentoring organization that initiated the treatment.

[8] An additional requirement was that the children speak German, although we did not exclude families with a migration background. In fact, 34.4% of the participating children have at least one parent who was not born in Germany.

[9] The distribution was calculated using the 2009 wave of the German Socio-Economic Panel (SOEP) (Wagner, Frick, and Schupp, 2007) and the cross-sectional weights provided therein.

[10] With respect to single-parent status, we expected that single-parent households dispose of less time resources to spend with their children than households with two parents. Using survey data (mother interviews), we can actually show this. Children living in single-parent households spend 45.8% more time "alone at home" than those in households with two parents.

[11] Given the larger relative supply of mentors in Bonn, we also assigned a higher share of children in Bonn to the ITT group. Therefore, assignment into treatment was random conditional on city of residence.

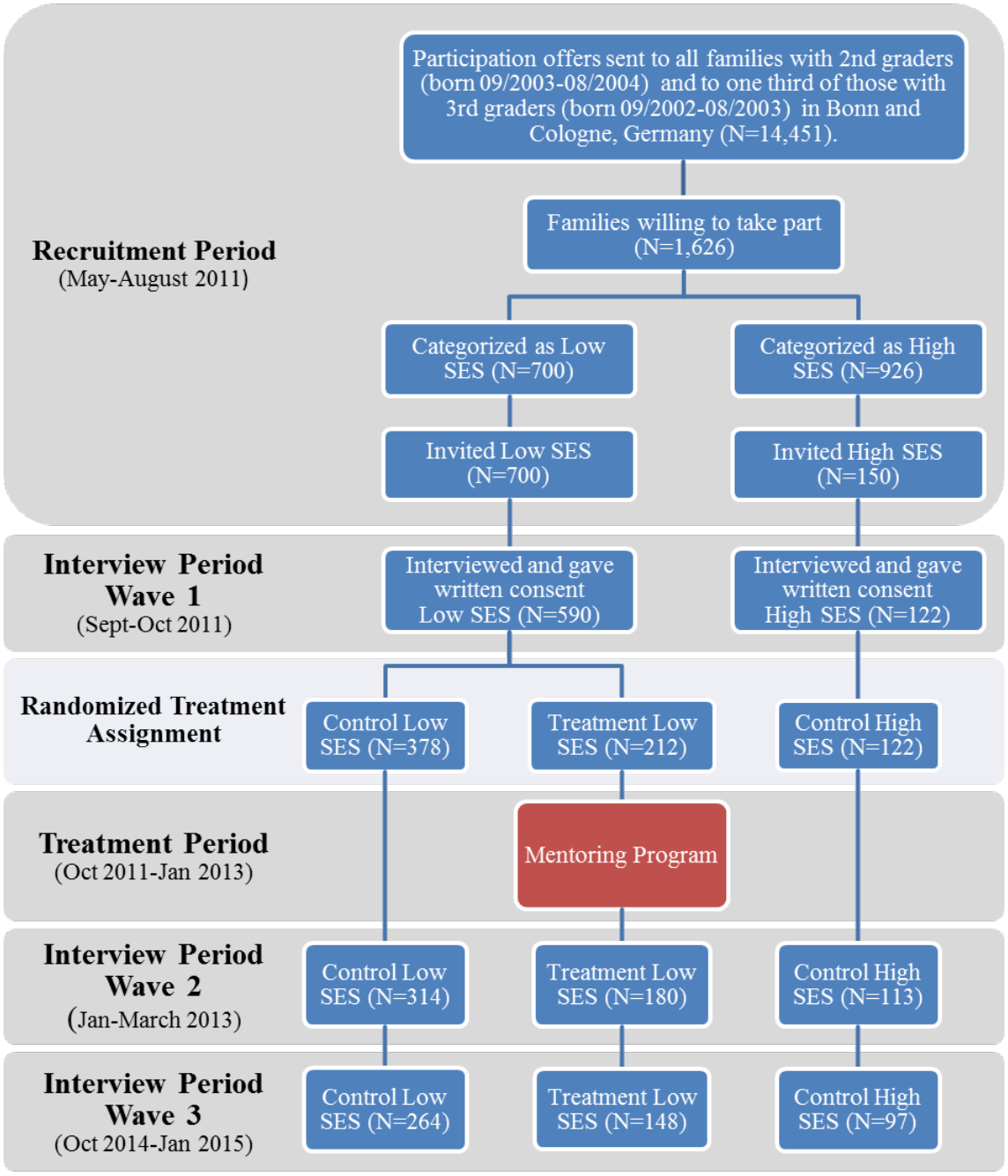

Figure 1: Flow chart of sampling and procedural details.

As a second control group, we also invited 150 randomly-chosen high SES families (among those who had answered the information letter and had given their written consent) of whom 122 took part in the wave 1 interviews (Control High SES).

After the one-year treatment period, all families who had participated in wave 1 were invited to take part in the post-treatment wave (wave 2). 85.3% of them (607 out of 712) took part in this second wave of interviews. This is our core sample. Two years after the end of the treatment period, all families who had participated

in the second wave were invited to take part in the third wave of interviews. 83.9% (509 out of 607) took part in this two-year follow-up.[12]

2.2 Setting of interviews and experiments

In all interviews, the child was accompanied by one parent. In 95% of the cases, the interviewed parent was the biological mother. For convenience, we therefore use the term "mother" for the adult who was interviewed. Overall, we collected three waves of data. In waves 1 and 2, interviews took place at central locations in either Bonn or Cologne. For this purpose, we rented two large flats, one in each city. The interviews and experiments were conducted by trained interviewers with a background in psychology or education science. In wave 3, interviews took place at participants' homes. In this wave, interviews and experiments were conducted by experienced and trained interviewers of the same professional surveying company that administers the SOEP (Wagner, Frick, and Schupp, 2007). The interviews and experiments were conducted according to a detailed protocol (see appendix section C), which was identical across all three waves. Overall, interviews lasted about one hour. For participation in the interview, mothers received 35 Euros in wave 1 and 45 Euros in waves 2 and 3, respectively.

We took care to create a pleasant, non-stressful interview situation by seating a mother and her child in the same room. However, in order to avoid interaction between the two, a standardized seating plan ensured that mother and child could not directly see each other. In addition, they were not allowed to communicate. One experimenter ran experiments with only one child at a time. During the experiments, mothers completed a comprehensive survey covering topics such as basic information about the child, mother assessments of personality and attitudes of the child, socio-economic background of the family, details on how parents spend time with the child including joint activities, as well as economic and social preferences of the mother.

The experiments run with children were incentivized using an experimental currency called "stars". At the end of the interview, children exchanged their stars for toys (in waves 1 and 2) or money (in wave 3). Toys were arranged in four categories that visibly increased in objective value and subjective attractiveness to children (see figure A2). During the experiments, children knew that more stars would result in the option to choose toys from a higher category.[13] In wave 3, children had

[12] See section 3.4 for a discussion of baseline balance and attrition.

[13] We ensured that each additional star that would not result in a higher category was nevertheless valuable: these stars were exchanged into "Lego" bricks.

reached an age where toys are no longer appropriate as an incentive. We therefore changed to money in this wave, with one "star" corresponding to 0.30 Euro.

In order to circumvent biased assessments and experimenter demand effects, we implemented "firewalls" between the mentoring program, data collection and research, including all interviews and experiments. They comprised several measures: (i) mentors received no information about the elicited measures, to avoid any form of "training to the test"; (ii) at no point in time were interviewers informed about the purpose of the study or the group assignment of the participating families (neither treatment/control nor high/low SES); (iii) the intervention was not mentioned during the data collection phase and (iv) the research team never interacted directly with the children or their parents.

2.3 Measures of prosociality and social interaction

In order to obtain a comprehensive assessment of child prosociality, we elicit prosocial motivation (altruism), beliefs about the prosociality of others (trust) and prosocial actions (other-regarding behaviors). Moreover, we extract information by means of data collection methods (choice experiments, survey measures) and using statements from children and mothers. The following describes each measure of child prosociality, our measures of prosociality in mothers and mentors, as well as our measure of social interaction patterns.

Altruism: Incentivized dictator game experiments

We measured children's altruism using three incentivized versions of dictator games with the participating children in the role of the decision-makers ("dictators"). The children had to allocate amounts of the experimental currency (stars) between him-/herself and another anonymously-matched child of the same age. In particular, we conducted one binary dictator game and two continuous versions of dictator games with varying receivers (see the appendix section C for experimental protocols).

In the binary dictator game (see Fehr, Bernhard, and Rockenbach, 2008), children had to decide between two possible allocations of two stars, regarding themselves and another unknown child from the same city. One option was that both, decision-maker and receiver, receive one star (1,1). The other option implied that the decision-maker received two stars while the receiver received no star (2,0). In both continuous versions, the decision-makers were endowed with six stars and could choose how to distribute the six stars between themselves and the other child, respectively. The two continuous versions differed by the characteristics of the receiver. In one version, the receiver was an unknown child who lives in a city nearby. In the

other version, the receiver was an unknown child who lives in an "African country" and who cannot live with his/her parents because they are either "poor or ill" or "have passed away".

In all three versions, the allocations were physically shown and the children answered control questions. Our measure of altruism is the average share of stars given in each of the three dictator games. In our wave 2 sample, the mean sharing rate was 0.390 with a standard deviation of 0.156 ($N = 606$; all children answered the control questions correctly and only one child needed a repetition of the rules, one missing observation).

Importantly, all decisions had real consequences for the participating children and the anonymous receivers. We cooperated with three charities (in Cologne, Bonn and Togo, respectively) to implement the allocation decisions as described. All children in the role of receivers (Cologne, Bonn and Togo) benefited from receiving stars in form of toys. In order to benefit receivers in Cologne and Bonn, we collaborated with two local charity organizations. To implement the dictator game outcomes with children who "live in Africa", we collaborated with an SOS Children's Village in Togo.

Trust: Questionnaire answers of the child

Children answered three questions concerning trust. These survey items are taken from the SOEP (Wagner, Frick, and Schupp, 2007) and have been experimentally validated (Fehr et al., 2002; Falk et al., 2016). We slightly adapted these items to make them appropriate for children in the age range under study. In particular, the statements read as follows: "One can trust other people", "Other people have good intentions towards me" and "One can rely on other people, even if one does not know them well". The statements were read out aloud by the interviewer and children indicated how strongly they agree with the statements using a five-point Likert scale ranging from "totally disagree" to "totally agree". As shown in figure A3, the scale was printed on an extra sheet of paper and additionally visualized. In order to further facilitate understanding, the interviewer explained the procedure using a simple neutral example item ("I like Spaghetti"). The average response to the three items is our measure of a child's trust. In our wave 2 sample, the mean is 3.193 with a standard deviation of 0.765 ($N = 607$; all 607 children answered all three trust questions).

Other-regarding behavior: Mother survey

As part of the mother survey, every mother assessed her child's other-regarding behavior in everyday life using the Prosocial Scale of the Strength and Difficulties

Questionnaire (SDQ) (Goodman, 1997). The SDQ is a well-established behavioral screening survey comprising five sub-scales. The Prosocial Scale includes five items, which read as follows:[14] My child is... "considerate of other people's feelings", "shares readily with other children", "helpful if someone is hurt, upset or feeling ill", "kind to younger children" and "often volunteers to help others". Responses were given on a seven-point Likert scale ranging from "does not apply at all" (1) to "applies completely" (7). The average answer to these five items is our measure of a child's other-regarding behavior in everyday life. In our wave 2 sample, the mean is 5.837 with a standard deviation of 0.972 ($N = 605$; according to the manual, we include an observation if at least three out of five items are completed; two missing values remain).

Joint measure: Prosociality

In sum, we obtained three critical facets of prosociality – altruism, trust and other-regarding behavior – combining experimental choice and survey data, and exploiting responses of the child and the mother. Each single measure is based on multiple responses (three dictator games, three trust questions and five other-regarding behavior items), which reduces measurement error. Collecting information from mothers and children bears the advantage that some of the measures are collected independently of the subject of interest, thus reducing demand and observability effects. For the main analysis, we collapse the three facets of the underlying prosocial disposition into one joint measure of prosociality. This measure is the equally-weighted score of the standardized measures of the three facets. In order to ease comparability, for all analyses in this paper (waves 1, 2 and 3) we use wave 2 means and standard deviations for standardization. To limit missing observations, if one of the three facets is missing the joint measure is calculated as the equally-weighted score of the two observed facets.

Prosociality of mothers and mentors

In order to obtain measures of prosociality for mothers and mentors, respectively, we proceed as similarly as possible as for the children: we construct an equally-weighted score using standardized measures of altruism, trust and other-regarding behavior. All measures are collected using established and validated survey items. Altruism is measured using the question "How would you assess your willingness to share with others without expecting anything in return, for example your willingness to give to charity?" (Falk et al., 2016) and the share of actual altruistic activities in everyday life (for details, see section B.2). Trust was measured using the two items

[14] We used the wording of the official German SDQ version, see www.sdqinfo.org.

"In general, one can trust people" (Fehr et al., 2002) and "As long as I am not convinced otherwise, I always assume that people have only the best intentions" (Falk et al., 2016). Responses were given on an eleven-point Likert scale. Other-regarding behavior was measured using the Big Five dimension Agreeableness in form of a three-item version (seven-point Likert scale) (Lang et al., 2011). The items read as follows: I see myself as someone who "is sometimes somewhat rude to others" (reversed), "has a forgiving nature" and "is considerate and kind to others". Note that the Agreeableness items use a similar wording as the Prosocial Scale of the SDQ, which we use to measure children's other-regarding behavior. Moreover, Agreeableness is theoretically and empirically related to concepts of other-regarding behavior (Becker et al., 2012). As for children, we construct an equally-weighted score of the three measures as our measure of prosociality for mothers and mentors.

Social interaction patterns
In order to elicit intense social parent-child interactions, we asked mothers how they spend time together with their child. We focused on joint activities similar to those that mentors and mentees engage in: having a conversation, having a snack together (e.g., a cake), playing board or card games, playing music together or going to music lessons. For each item, mothers were asked: "How many times during the last 14 days have you or the main caregiver done the following activities together with your child?". Our measure of the intensity of social interaction of mother and child is the average share of highly-interactive activities out of all activities per afternoon, with a mean of 42.2%. We collected very similar information from the mentors. In line with the design of the program laid out in the following section, the share of highly-interactive activities at mentor-mentee meetings was 94.7% and thus much higher (see also section B.5).

2.4 The mentoring intervention

The intervention that we implemented is a well-established non-profit mentoring program called "Balu und Du" (German for "Baloo and you"). In this program, elementary school children are provided with a mentor for the duration of one year. The mentors, called Baloos, are mainly university students (aged from 18 to 30) who voluntarily care for their mentees, called Mowglis. The conceptual idea of the program is to focus on "informal learning", a concept that integrates and reinforces learning processes in children's everyday life. According to informal learning, the

mentors act as role models and "benevolent friends" who encourage the acquisition of new ideas and skills by enriching the social environment of the children.[15]

On a practical level, a mentored child typically spends one afternoon per week in one-to-one interaction with his/her mentor. During this time, they engage in joint activities, which are adapted to the individual needs, strengths, weaknesses and interests of the child (and mentor). Examples include visiting a zoo, museum or playground, cooking, doing handicraft, ice skating or simply having a conversation. Hence, a child spends time with an additional attachment figure and role model, gains new experiences and learns that he/she is valued based on the fact that the mentor regularly spends time with him/her.

The mentoring program is embedded in a professional structure. On a weekly basis, mentors complete an online diary in which they report the activities that they have engaged in, as well as potential problems of the mentor-child relationship. Program coordinators read and comment on these diaries, and provide support. These coordinators are trained and paid professionals in education science or psychology and they provide supervision and advice to mentors. They also organize bi-weekly monitoring meetings where mentors receive suggestions for activities with the mentored child and discuss potential problems. To date, the "Baloo and You" mentoring program has arranged and supervised around 10,000 mentor-child relationships in more than 50 different locations in Germany.[16]

The mentoring program is designed to last up to 12 months. In our sample, the actual average duration of mentor-mentee relations was 9.3 months (figure A1). Variation in duration is mainly due to unforeseeable events such as moving decisions of parents or mentors due to a job change. On average, treated children met their mentor 22.8 times (std. dev. 11.9), typically for a whole afternoon (amounting to a total average of around 92 hours).

We transmitted household addresses of all randomly-selected families to the mentoring organization. The actual matching process of mentor and mentee is part of the program and was conducted by the organization. Each child in the ITT group could potentially be matched, but not all selected children were effectively matched with a mentor. A mentor-mentee match was successfully implemented for 74% of the ITT group children. For 26% of the children, matches could not be realized due to a local shortage of mentors, mentor refusals or coordination problems between

[15]For further details, see Müller-Kohlenberg and Drexler (2013).

[16]The program has been honored with several public awards, e.g., by the Robert Bosch Foundation in 2011 and the federal government of North Rhine-Westphalia (Germany) in 2006. More details about the mentoring program can be found on www.balu-und-du.de and in an overview article by Müller-Kohlenberg and Drexler (2013).

mentors and families (e.g., pregnancy of the mentor, moving of mentor or family, etc.). Most of these children were never contacted by the organization. In the main analysis, we focus on ITT effects and discuss local average treatment effects (LATE) in section 3.4.

3 Results

In this section, we first provide descriptive evidence on the relationship between parental SES, parental inputs and child prosociality among the Control High SES and the Control Low SES group children. Section 3.2 then unveils the causal effect of enriching a child's social environment on prosociality and section 3.3 presents a mediation analysis concerning the underlying mechanisms and evinces how prosociality develops in elementary school children. Finally, section 3.4 provides numerous robustness checks. This includes tests for baseline balance, the absence of selective attrition and a confirmation of our main results using representative weights and difference-in-differences estimations.

For most analyses, we use the wave 2 sample ($N = 607$; 47.0% girls; age at the start of the program: Mean = 7.76 years, std. dev. = 0.48). However, in order to study long-run effects of the intervention and persistence of parental background over a two-year developmental period, we also refer to wave 3 data.

3.1 SES and parental input: descriptive evidence

The few papers relating child prosociality to family background characteristics suggest that children from high SES families are more prosocial than those from low SES backgrounds (Bauer, Chytilová, and Pertold-Gebicka, 2014; Benenson, Pascoe, and Radmore, 2007). As a first step, we thus explore systematic differences in prosociality among children who grow up in high and low SES families. Table 1 reports differences in prosociality for the two untreated groups of children, Control Low SES and Control High SES. A regression of child prosociality on a dummy variable for High SES (see column (1)) reveals that children from high SES households are substantially and significantly more prosocial than those from low SES households ($p < 0.05$). In terms of effect size, the difference amounts to 22.6% of a standard deviation.[17] This finding of lower levels of prosociality in low SES children in comparison to high SES children also holds if we consider alternative definitions of SES.

[17]This gap is comparable to the SES difference that we observe for a representative sample of adult individuals (see section B.3).

	Prosociality			
	(1)	(2)	(3)	(4)
High SES	0.226**			0.140
	[0.106]			[0.105]
Prosociality of Mother		0.258***		0.228***
		[0.048]		[0.048]
Social Interaction			0.214***	0.167**
			[0.048]	[0.049]
Constant	-0.127**	-0.083*	-0.064	-0.117**
	[0.059]	[0.048]	[0.048]	[0.056]
Observations	418	418	418	418
R^2	0.010	0.067	0.042	0.098

Table 1: Prosociality and parental background. Prosociality measures of mother and child are constructed as equally-weighted scores of standardized measures of altruism, trust and other-regarding behavior, respectively. Prosociality and social interaction measures are standardized, High SES is a dummy. For reasons of comparability, the sample is restricted to individuals with non-missing values in all used variables. Data come from wave 2. Coefficients are OLS estimates with White robust standard errors in brackets. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

In figure A4, we explore the SES gaps for all possible combinations of our low SES criteria (low income, low education and single-parent status). For each combination, we find that children from low SES families score lower on prosociality than children from high SES families.[18]

What determines the socio-economic disparities in child prosociality? Informed by the literature in developmental psychology, we identify and study two important drivers: prosocial attachment figures and intense social interactions. Prosocial attachment figures enable children to observe and imitate a prosocial role model and thereby internalize prosocial beliefs, preferences and behaviors (Skinner, 1953; Williamson, Donohue, and Tully, 2013). Moreover, intense social interactions promote the interplay between action and reaction as an important feedback provider that is essential to reinforce prosocial attitudes (Eisenberg, Fabes, and Spinrad, 2006; Eisenberg, Spinrad, and Knafo-Noam, 2015).

Our data display important SES-related differences in these stimuli. Children in high SES live with adults who are themselves more prosocial. On average, high

[18]Another potentially important difference in family background concerns migration status. 34.4% of the children in our sample have a migration background, defined as having a mother and/or father born outside of Germany. 23.4% of the children with a migration background have Turkish roots and represent the largest group among the immigrants. The data indicate that neither general migration background ($p = 0.729, N = 418$, two-sided t-test) nor specifically Turkish background is related to children's prosociality ($p = 0.609, N = 418$, two-sided t-test).

SES mothers score 15.6% of a standard deviation higher on prosociality than low SES mothers in our data ($p = 0.160, N = 418$, two-sided t-test). Moreover, high SES families spend on average 30.5% of a standard deviation more time on socially-interactive activities compared with low SES families ($p < 0.01, N = 418$, two-sided t-test).

As conjectured, these differences in maternal prosociality and interaction patterns also predict child outcomes. Column (2) in table 1 indicates that mothers' prosocial attitudes are strongly associated with children's prosociality. A one standard deviation increase in mother's prosociality is related to a 25.8% of a standard deviation increase in her child's prosociality ($p < 0.01$).[19] Moreover, column (3) shows that spending more time on intense social interactions positively relates to children's prosociality, as children who experience a one standard deviation higher intensity of social interaction with their mothers are 21.4% of a standard deviation more prosocial ($p < 0.01$).

In column (4) in table 1, we regress children's prosociality jointly on SES, social interaction and mothers' prosociality. While the coefficients of social interaction and prosociality of the mother remain significant, the coefficient of SES drops by 38% and is no longer significantly different from zero. These conditional correlations are in line with the interpretation that it is not SES per se but rather alterable factors in the child's household environment that shape a child's prosociality. Following this line of thought, subsequent analyses will answer two research questions: (i) Is prosociality malleable in elementary school children and can it be enhanced by enriching a child's social environment? (ii) What is the respective importance of different inputs such as intense social interactions and prosocial attachment figures for the formation of prosociality?

The remainder of this paper exploits intervention-induced variation in the children's social environment to provide causal evidence on the malleability of prosociality in children. It then dissects how prosociality develops in elementary school children and how it can be enhanced.

3.2 The causal effect of social environment on prosociality

The correlational evidence reported in the previous section suggests that prosociality in children forms in response to the quality of their social environment. The randomized controlled implementation of the intervention allows us to study whether

[19] This is also in line with the literature on intergenerational transmission of economic preferences, see Bisin and Verdier (2001), Dohmen et al. (2012), Kosse and Pfeiffer (2012), Alan et al. (2017), Zumbuehl, Dohmen, and Pfann (2013).

this relationship is causal. Children participating in the mentoring program are randomly assigned to an enriched social environment in form of experiencing an additional prosocial attachment figure and more intense social interactions. As shown above, both of these experiences are relatively scarce in low SES families compared with high SES families.

The left panel of figure 2 shows our first main result. In wave 2 (post-treatment), the Treatment Low SES children score 27.3% of a standard deviation higher on the prosociality measure than those from the Control Low SES group ($p < 0.01, N = 494$, two-sided t-test). Figure 2 also indicates that the high-low SES developmental gap in prosociality is closed for children in the treatment group: children from Treatment Low SES and Control High SES score very similarly on prosociality ($p = 0.651, N = 293$, two-sided t-test). The positive effect of the intervention is not only observed for the joint measure of prosociality but also holds independently for the three facets of prosociality, altruism ($p < 0.05$), trust ($p = 0.05$) and other-regarding behavior ($p = 0.266$) (see table A1).[20] Our measure of other-regarding behavior might be least affected by the treatment, because parent-reported measures tend to be noisier and less accurate than, e.g., experimental measures that are based on revealed-preferences especially in disadvantaged samples (Borghans et al., 2008; Sandner and Jungmann, 2016). In fact, if this sub-scale is omitted from the overall prosociality score, our results increase in size and significance (see column (5) in table A1).

Recent literature suggests that initially-disadvantaged children tend to profit most from childhood interventions (Duncan and Magnuson, 2013; Heckman and Mosso, 2014; Heckman and García, 2017). It is thus conceivable that children who experience low levels of prosociality-promoting parental inputs benefit relatively more from an enriched social environment. This is exactly what we find. As documented in table A2, the treatment is most effective among children whose mothers score low on prosociality. In a regression of child prosociality on maternal prosociality, the treatment dummy and its interaction, the respective interaction coefficient is significant ($p < 0.05$), indicating that the treatment effect increases by 21.4% of a standard deviation of prosociality if the mother scores one standard deviation lower on prosociality. A similar effect is found for the interaction of treatment and the intensity of mother-child interactions. The treatment effect increases by 19.2% of a standard deviation of prosociality ($p < 0.05$) if the child experiences one standard deviation less intensive social interaction in his/her home environment.

[20] A joint F-test rejects the null-hypothesis of a zero treatment effect for the three facets of prosociality (altruism, trust, other-regarding behavior; $p = 0.010, F = 3.78, N = 491$).

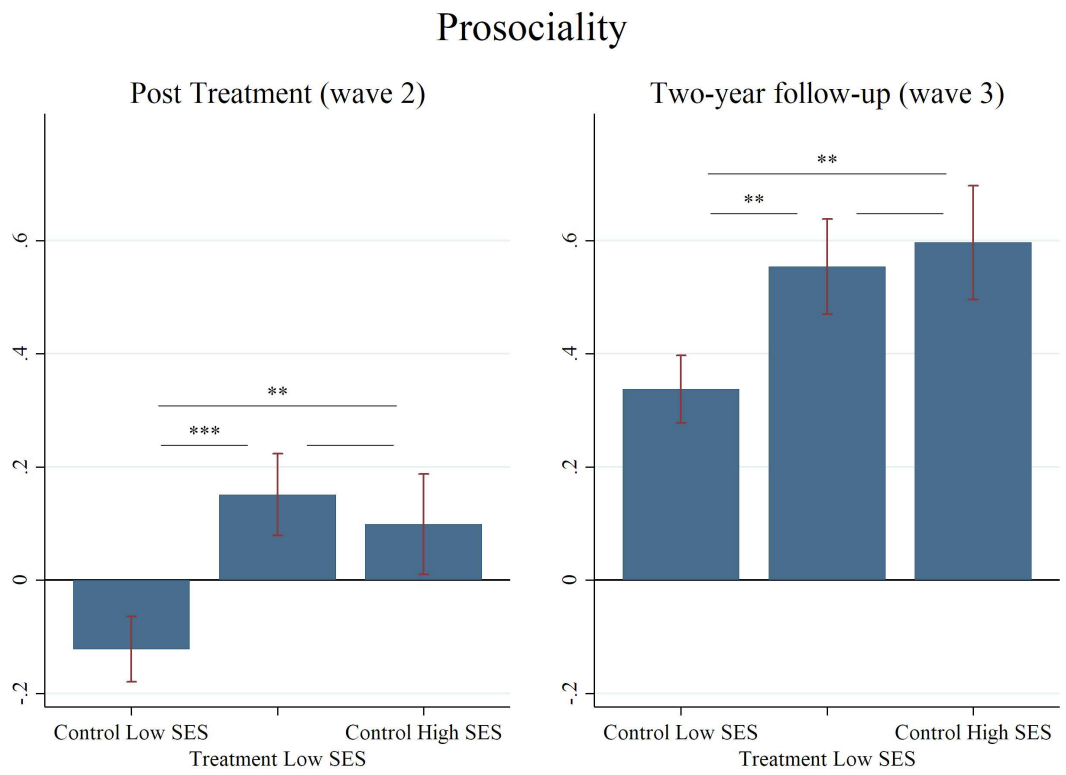

Figure 2: Significantly higher levels of prosociality for treated children compared to untreated children (Treatment Low SES vs. Control Low SES). There is no significant difference between Treatment Low SES and Control High SES. The scale on the y-axis indicates z-scores (i.e., standardized measures) of children's prosociality. Standardization is conducted using the distribution of wave 2. Error bars show standard errors of the means. *** and ** indicate significant differences at the 1 and 5 percent level, respectively (two-sided t-tests). Left panel: N(Treatment Low SES vs. Control Low SES) = 494, N(Control High SES vs. Control Low SES) = 427, N(Treatment Low SES vs. Control High SES) = 293), right panel: N(Treatment Low SES vs. Control Low SES) = 411, N(Control High SES vs. Control Low SES) = 358, N(Treatment Low SES vs. Control High SES) = 243).

We now turn to the important question of whether the observed effects of an enriched social environment on child prosociality are enduring. Even short and transitory changes in personality may be crucial, e.g., if they occur during critical transition periods. However, finding lasting effects with respect to temporary changes in the social environment would strengthen the relevance and credibility of our findings, as well as underscoring the potential of childhood interventions in general. In order to investigate this issue, we use the two-year follow-up data (wave 3), see figure 2 (right panel).

The figure illustrates three important findings. First, over the time span of two years, we observe a general increase in prosociality in our sample of elementary school children. The average increase amounts to 43.8% of a wave 2 standard deviation and

holds for all groups under study, i.e., irrespective of SES and treatment.[21] To the best of our knowledge, this is the first documentation of an increase in prosociality among children, complementing and supporting previous studies using cross-sectional data (see e.g., Sutter and Kocher, 2007; Fehr, Bernhard, and Rockenbach, 2008; Fehr, Glätzle-Rützler, and Sutter, 2013; Bauer, Chytilová, and Pertold-Gebicka, 2014). Second, the high-low SES developmental gap in prosociality that we have seen in wave 2 persists to wave 3. The difference in prosociality between Control Low SES and Control High SES in wave 3 is similar to the one in wave 2 and adds up to a 25.9% of a wave 2 standard deviation ($p < 0.05, N = 358$, two-sided t-test).[22] Third, and most importantly, the treatment effect reported in the left panel of figure 2 proves to be remarkably robust over time. Two years after the end of the intervention, treated children display significantly higher levels of prosociality than those from the control group (21.7% of a wave 2 standard deviation, $p < 0.05, N = 411$, two-sided t-test). Consequently, prosociality in wave 3 does not significantly differ between Treatment Low SES and Control High SES children ($p = 0.749, N = 243$, two-sided t-test). In other words, the developmental gap that was closed in response to treatment remains closed more than two years after the intervention.

The above results substantiate the notion that prosociality is malleable through an enrichment of children's social environment. They also indicate that children who receive fewer prosociality-promoting parental inputs benefit relatively more from the intervention. The subsequent section expands on these findings by seeking to quantify the importance of intense social interactions and prosocial attachment figures for the formation of child prosociality.

3.3 Mechanisms

In this sub-section, we present a production function of child prosociality that reconciles the evidence of the previous two subsections and forms the basis for a mediation analysis. Our aim is to relate the general process of prosociality formation to the mentoring program's main components and quantify the relative importance of prosocial attachment figures and intense social interactions.

[21] A test for differences in post-treatment prosociality growth rates between the Treatment Low SES and Control Low SES groups reveals that these differences are small and not significantly different from zero ($p = 0.611$).

[22] Among current adult cohorts in the SOEP, the SES gap in prosociality amounts to 0.18 of a standard deviation, suggesting that differences in prosocial behavior are likely to be persistent and do not fade out as individuals grow older (see table B2 in section B.3).

Following the approach and notation laid out in Heckman, Pinto, and Savelyev (2013), we focus on a linear production function and write child i's prosociality, when the program assignment is set to "treated" ($d = 1$) or "control" ($d = 0$), as:

$$P_{i,d}^{C} = \kappa_d + \alpha_d^P P_{i,d}^A + \alpha_d^I I_{i,d}^A + \alpha_d^U U_{i,d} + \beta_d X_i + \epsilon_{i,d}, \quad d \in \{0, 1\}, \tag{1}$$

where $P_{i,d}^{C}$ represents counterfactual prosociality of child i. κ_d is an intercept and α_d^P and α_d^I are scalar parameters, denoting the effects of prosocial attachment figures and social interactions on child prosociality. Moreover, α_d^U is a vector denoting the effect of several unobserved factors ($U_{i,d}$) on child prosociality, β_d is a vector that captures the effect of pre-program variables, such as the child's region of residence. Finally, ϵ_i denotes an error term that is independent of the mechanisms and predetermined variables. The above equation thus reflects that attachment figure prosociality and intense social interactions are important determinants of child prosociality.

Assessing the relative importance of prosocial attachment figures and intense social interactions for the program's success requires measurements of these variables for treated and control group children. Control group children never see a mentor and are therefore mainly affected by their mother's prosociality and regular social interactions at home. For these children, and for the ITT children who never met a mentor, P^A stands for maternal prosociality[23] ($P^A = P^M$) and I^A is measured as the average (weekly) share of highly-interactive activities out of all activities experienced per afternoon. Children in the treatment group who interacted with a mentor are influenced by mother *and* mentor. The relative importance of mother and mentor as attachment figures is not self-evident. Yet, the child's attachment and degree of imitation likely depends on the quality of relationship with either of them.[24] Hence, we construct relation-specific weights ω. These weights are derived from survey items designed to measure the quality of the respective relationship between the child and his/her mother (M) or mentor (B), respectively (see section B.5 for details).[25] P^A for treated children is thus given by $P^A = \omega \cdot P^M + (1 - \omega) \cdot P^B$. Social interactions ($I^A$) are measured as the average share of highly-interactive

[23]Prosociality was only measured in mothers. However, e.g., Dohmen et al. (2012) document a strong positive correlation of preferences within couples, which is consistent with positive assortative mating. Positive assortative mating based on preferences is further predicted by the models of Bisin and Verdier (2000, 2001) on the cultural transmission of preferences.

[24]For a discussion on the importance of the quality of relationship, see MacCallum and Beltman (2002).

[25]An alternative possibility would be to choose ad-hoc weights with $\omega \in (0, 1)$. In figure A7, we show robustness results using various ad-hoc weights. If we choose equal weights ($\omega = 0.5$), our analysis yields similar results. For, e.g., $\omega = 0.4$ ($\omega = 0.6$), the prosocial attachment figure slightly increases (decreases) in importance.

activities per afternoon, taking into account that treated children spend up to one afternoon per week with their mentors. There exits substantial heterogeneity in the share of highly-interactive activities conducted with both mothers and mentors, as documented in the respective time-use questionnaires (see section B.5).

Random assignment into treatment and control groups ensures that the ITT effect on our mechanism variables is easily computed as a mean difference. The results displayed in figure A5 show that children in the treatment group are exposed to attachment figures that are on average 55% of a standard deviation more prosocial ($p < 0.01$, two-sided t-test). The reason is that the self-selected volunteer mentors are on average more than a standard deviation more prosocial than the low SES mothers in our sample (see table B4 and section B.6). Figure A5 also shows that the ITT children engage on average in 14% more intense social activities per afternoon (40% in terms of standard deviations, $p < 0.01$, two-sided t-test). The mentoring intervention is thus very intensive in terms of providing these two inputs. Most notably, investments in the Treatment Low SES group are even higher than in the Control High SES group. Taking these findings together, the mentoring program likely benefits low SES children by providing a substantial amount of resources and stimuli that are both critical and scarce in the given family environment.

A straightforward decomposition of the program's ITT effect on child prosociality can be achieved under two additional assumptions. First, we conjecture that the impact of the mechanism and control variables on child prosociality is the same in the two groups, such that $\alpha_1^P = \alpha_0^P$, $\alpha_1^I = \alpha_0^I$ and $\beta_1 = \beta_0$.[26] Second, we assume that program-induced increments in measured (P^A, I^A) and unmeasured (U) mechanism variables are statistically independent conditional on X and D. Heckman, Pinto, and Savelyev (2013) show that under these assumptions the overall prosociality effect can be decomposed into:

$$\underbrace{\mathbb{E}[P_{i,1}^C - P_{i,0}^C]}_{\text{prosociality effect}} = \underbrace{\tau_1 - \tau_0}_{\text{unmeasured}} + \underbrace{\alpha^P \ \mathbb{E}[P_{i,1}^A - P_{i,0}^A]}_{\text{prosocial attachment}} + \underbrace{\alpha^I \ \mathbb{E}[I_{i,1}^A - I_{i,0}^A]}_{\text{intense social interactions}} \quad , \tag{2}$$

where $\tau = \kappa_d + \sum_{j \in J_U} \alpha_d^j \ \mathbb{E}[U_d^j]$, such that $\tau_1 - \tau_0$ captures the contribution of treatment-induced changes in a number of J_U unmeasured mechanism variables.[27]

[26] Following Heckman, Pinto, and Savelyev (2013), we empirically test these assumptions in a joint model with interaction effects. The assumptions are not rejected at any conventional level, neither individually nor jointly. Relaxing the assumption that the parameters of the production function are the same in treatment and control groups gives a standard Oaxaca-Blinder decomposition yielding similar results.

[27] The above approach relies on linearity of the production function. More general non- and semi-parametric identification of the ITT effect has been discussed in Yamamoto (2013) and Keele, Tin-

Figure 3 displays our estimated decomposition of the overall program effect into exogenously induced changes in the exposure to prosocial attachment figures, changes in intense social interactions and other factors.[28]

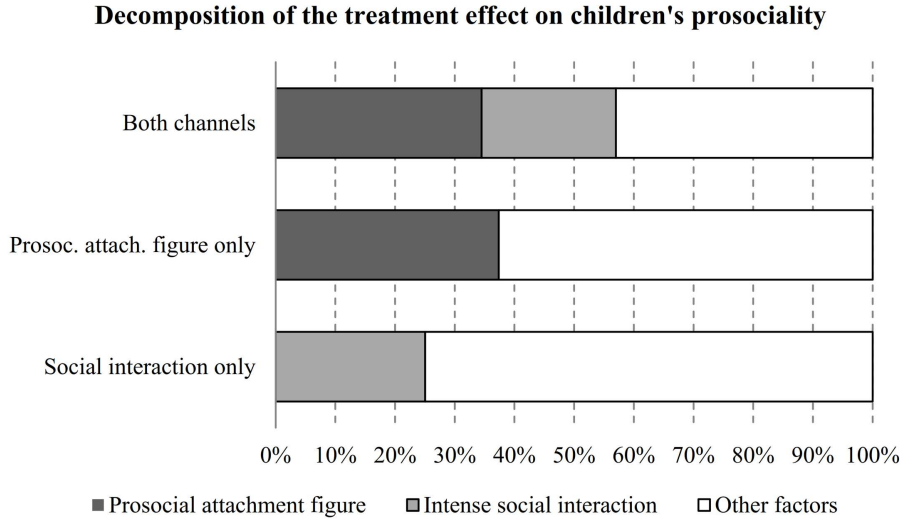

Figure 3: The figure shows the estimates of the decomposition as described in equation 2. Each bar represents the total treatment effect normalized to 100%. The dark gray area displays the share of the treatment effect explained by changes in prosociality of attachment figures. The light gray area displays the share of the treatment effect explained by the increase in intense social interactions. The white area indicates unmeasured other factors. The bar in line 1 shows the results for decomposition into both observed channels (prosocial attachment figures and intense social interactions). The middle bar shows the results for the attachment figure channel only. The bottom bar shows the results for the social interaction channel only. Two-sided p-values: $p < 0.01$ for all channels in all three decompositions. For details on the decomposition, see Heckman, Pinto, and Savelyev (2013).

The bar in line 1 shows that the program's main components together can explain about 60% of the change in child prosociality. Moreover, roughly two-thirds of the explained ITT effect can be ascribed to the increase in prosociality of attachment figures and one-third to more intense social interactions.[29] The bars in the middle and bottom parts of the figure show the results for both inputs separately. They illustrate how the results would change if one of the inputs was ignored and ascribed to the unmeasured component in equation (2) instead. The explained shares

gley, and Yamamoto (2015). For identification that relies on instrumental variables for treatment compliance and mediators, see Frölich and Huber (2017).

[28]Note that all results are invariant to linear transformations of the measures.

[29]If we conduct the same decomposition to explain the SES gap documented in section 3.1, we find that differences in maternal prosociality and intense social interactions together explain about 38% of the SES difference in child prosociality, with both factors being of roughly equal importance.

remain similar, indicating that both inputs affect child prosociality separately and independently.

3.4 Robustness checks and alternative treatment estimates

The aim of this section is to assess the external validity and robustness of our results as well as alternative treatment estimates. For this purpose, we first discuss results based on representative population weights and weights that account for potential selective attrition. We then present evidence from difference-in-difference and two-stage least squares estimators. Finally, we assess our results in light of multiple hypothesis testing and assess joint characteristics that improve the mentor-mentee match quality.

The families in our sample are not necessarily representative of the German population, given that they declared interest in the mentoring program and participation in our study. For example, it is conceivable that mothers who take an interest in the program on behalf of their children are themselves more prosocial or see the need for a mentor who spends interactive time with their child. Moreover, SES was part of the sampling scheme, such that families likely differ from a representative sample in terms of household income, parental education and single parenthood. We thus construct representative population weights based on maternal prosociality, intensity of mother-child interaction and different dimensions of SES. This is possible because a substantial part of the mother questionnaire matched the questionnaire of the SOEP, a dataset that is representative of the German population. The weights consider all family-related variables used in the main analysis: the three dimensions of SES, maternal prosociality and the intensity of mother-child interaction (see section B.4.1, for details).[30] The re-weighted ITT effects and SES gaps are very similar to the estimated main effects (see figure A6), except for a slight increase in the ITT effect (see table A3 for a comparison).

A further potential limitation for the interpretation of randomized intervention studies is selective attrition. If systematically correlated with treatment status or prosociality, attrition might bias our estimates. Since we have collected the outcome of interest (prosociality) not only after but also before treatment assignment, we can use these data to show that the lost to follow-up rates in waves 2 and 3 are neither related to pre-treatment prosociality, treatment assignment nor their interaction (see table A4). Moreover, re-weighting the observed data using inverse probabilities of participation in the wave 2 and 3 interviews confirms our findings reported in

[30]For a discussion on using weights based on propensity scores to assess the generalizability of results from randomized trials see, e.g., Stuart et al. (2011).

section 3.2 (see table A5).[31] In addition, we estimate treatment effects using the trimming procedure suggested by Lee (2009). Instead of correcting point estimates, this approach yields interval estimates of effect sizes based on extreme assumptions about selection. The results shown in table A6 are in line with the previously-reported results.

The pre-treatment measure of prosociality can further be used to investigate baseline imbalance. The randomization procedure was successful as pre-treatment prosociality does not differ by treatment status ($p = 0.662, N = 590$, two-sided t-test). Moreover, we can use the pre-treatment measure of prosociality to obtain difference-in-difference estimates as a robustness check of our main estimates. In sum, and consistent with baseline balance and non-selectivity of attrition, the diff-in-diff estimates (reported in table A7) confirm our main findings.

Not all children who were assigned to the treatment groups have actually taken part in the mentoring program. For 26%, the mentor-mentee match was either not initiated due to a shortage of mentors or matches could not be realized due to moving, other coordination problems or refusals by mentors or families.[32] We thus focus on ITT estimates since they only rely on random treatment assignment rather than specific assumptions about the mentor-mentee matching process implemented by the mentoring organization. The ITT is thus the policy-relevant effect of interest in the sense that it constitutes the average effect of informal mentoring during elementary school if offered to interested low SES families on a voluntary basis. Nevertheless, since the mentoring program takes effect through successful participation, we also provide information on local average treatment effects (LATE).[33] In table A8, we present two-stage least squares (2SLS) estimates using the random assignment as an instrument for actual treatment. Because, by design, control group children could not participate in the program, the LATE equals the average treatment effect on the treated, amounting to 39.8% of a standard deviation. Thus, in line with a matching

[31] The predicted probabilities result from a Probit model of a binary selection indicator (indicating whether or not wave 2 or 3 interviews were conducted) as a function of pre-treatment prosociality and treatment assignment.

[32] The complier children do not significantly differ from the non-compliers in any relevant dimension, such as initial child prosociality, mother prosociality, parental investments, etc. Participant selection on program gains for the majority of non-compliers is thus unlikely. In line with this presumption, the organization informed us that most initiated but uncompleted matches occurred because mentors moved away, were pregnant or changed their plans for other reasons.

[33] The LATE is informative about the average effect among the compliers. It answers the somewhat more difficult policy question: "What is the average effect of informal mentoring during elementary school in a group of interested low SES children who were offered a mentor and actually participated in the mentoring program?"

rate (compliance rate) of 73.9%, the LATE effect exceeds the ITT effect by about 35%.

Thus far, we have focused on testing our initial hypothesis that child prosociality can be enhanced by enriching a child's social environment. Throughout the paper, we have shown that the program affects prosociality by intense social interactions with a role model who strongly differs from the child's main caregiver in terms of prosociality. Given these findings, one might also expect to find treatment effects of the program on other economically-relevant preferences. We test this presumption for time and risk preferences, since the data collection project as a whole focused on fundamental economic preferences in children (see, Deckers et al., 2017). According to the role model hypothesis, however, one would expect to see a treatment effect along these other preference dimensions if mentors differed from the child's regular caregivers in terms of time and risk preferences. In table B4, we show that this presumption does not hold, mothers and mentors are hardly distinguishable with respect to patience and risk-taking.[34] Therefore, it is not surprising that we do not observe treatment effects along these other preference dimensions (for details, see section B.7). Nevertheless, it is comforting that the treatment effect on prosociality is still significantly different from zero at the 1 percent level (see table B5 for adjusted p-values) even after we conduct a multiple hypotheses correction, using the procedure outlined in Romano and Wolf (2005).[35]

As shown above, the increase in child prosociality tends to be highest if mentors and regular caregivers strongly differ in terms of prosociality and social interactions with the child. Moreover, the literature also suggests that one's social identification with a role model increases if there are core commonalities regarding gender and social background (Carrell, Page, and West, 2010; Kofoed and McGovney, 2017). Therefore, we explore whether similarities regarding gender, place of residence and parental background lead to more successful mentor-mentee matches. When matches were formed, only the gender of child and mentor and the addresses of families and mentors were known to "Balu und Du". Male mentors were only matched to male mentees and both mentors and mentees needed to live in the same city. In the following analysis, we condition on city and gender of the child and then exploit natural variation among matches. We evaluate successful matches by the subjective quality of the mentor-mentee relationship (as rated by the mentor, see also section B.5) and the increase in children's prosociality (between wave 1 and 2). The results are displayed in table A9. Regarding the gender composition, we find that same sex

[34]Further, note that intensity of social interaction as described in section 2.3 is also not significantly related to risk-taking ($p = 0.346$) or patience ($p = 0.753$), respectively.

[35]See Heckman et al. (2010a) and Romano and Wolf (2016) for a portrayal of the algorithm.

matches[36] tend to result in a better relationship quality ($p < 0.1$, two-sided t-test) and a slightly larger increase in children's prosociality ($p = 0.507$, two-sided t-test). We also find that matches are more beneficial if mentors and mentees live nearby in terms of commuting distance ($p = 0.632$ for relationship quality and $p < 0.05$ for the increase in prosociality, two-sided t-test). We do not observe similar effects for commonalities in parental background.

4 Conclusion

This study provides several important insights for understanding the formation of prosociality. First, we document the role of the social environment in terms of SES, social interaction patterns and maternal prosociality. Second, our panel data allow us to document a general within-subject increase in prosociality for elementary school-age children. Using the panel dimension, we also show that the effect of parental background is persistent over time. Third, our main result provides causal evidence on the effect of social environment on prosocial attitudes. The effect is significant both statistically and economically, and remarkably robust over time: the positive treatment effect of the enriched social environment on prosociality is enduring and observed two years after the end of the intervention. It is based on a comprehensive measure of prosociality, consisting of the three facets altruism, trust and general other-regarding behavior, using different elicitation methods and statements from different sources. Fourth, we provide evidence that prosocial attachment figures and intense social interactions are important drivers for the formation of child prosociality. Our study thus depicts both how prosociality forms and how it can be enhanced.

Our findings hold broad significance. Prosociality pervades human societies and is of fundamental importance at all levels of social interaction. Our results indicate that prosociality is malleable and provide insights concerning the effectiveness of early childhood interventions. Investments such as the mentoring program under study have the potential to systematically affect character formation with possible long-run benefits. Assuming a persistent program effect of 20% of a standard deviation in prosociality and using the estimates for gross wages displayed in table B1 yields, for example, a benefit-cost ratio of 3.84 and an internal rate of return of 8.5%.[37] However, the overall gains of the program are probably much larger,

[36] Given the constraint that only male mentees were matched to opposite or same sex mentors, identification comes from male mentees.

[37] The program benefits in terms of lifetime earnings can be approximated by multiplying the standardized gain in hourly gross wages (0.847, see table B1) with the treatment effect on the

because prosociality also impacts various other life domains, some of which are difficult to quantify (e.g., cooperative behavior, life satisfaction, well-being, antisocial behaviors, conduct problems or health).[38] In addition to these positive long-run effects, investments of the type that we study in this paper may be socially desirable because they close developmental gaps arising from socio-economic disparities in investments. This is important in light of increasing social inequalities and the intergenerational persistence of life outcomes (Case, Lubotsky, and Paxson, 2002; Currie and Moretti, 2003; Aizer and Currie, 2014; Putnam, 2015).

Our results also deliver insights regarding the effective implementation of programs targeted at the socio-emotional development of children. They suggest that mentoring works if the scope of activities and type of role models support its goals. Benefits are largest if the program leads to differences in the intensity and type of activities that a child experiences and if the mentor's characteristics set him/her apart from the child's usual caregivers. Policy-wise, these are important findings for several reasons. First, they imply that the effectiveness of any mentoring program hinges on the type of mentors who conduct it and that the selection of mentors is crucial. Second, our findings suggest that mentoring is particularly effective if it targets children who experience few prosocial attachment figures and intense social interactions, and if those children encounter mentorship experiences that stand out along these dimensions. To the extent that volunteer mentors tend to score high on social skills and prosocial motivations, our results thus also provide an explanation for the findings of other mentoring evaluations (Tierney and Grossman, 2000; Moodie and Fisher, 2009), e.g., regarding the reduction of antisocial behaviors, conduct problems or peer relationship difficulties. Last, our results point towards an improved match quality if mentors and mentees are of the same sex and if they live close-by. These findings are in line with the role model literature suggesting that the relationship quality between mentors and mentees improves if they share some common characteristics (Carrell, Page, and West, 2010; Kofoed and McGovney, 2017). Thus, if the objective function is to maximize treatment effects, the matching of mentors and mentees might be decisive for the overall success of the program. In line with the literature on social mobility and child resilience (see, e.g., Werner,

level of prosociality (0.2), the average number of yearly hours worked in Germany (about 1400, see, OECD, 2017) and the (predicted) average years of a working life (43). Lifetime earnings were discounted with a real discount rate of 3% (see, United States General Accounting Office, 1991; Heckman, Pinto, and Savelyev, 2013). For further details on the underlying assumptions and alternative estimates, see section B.8.

[38] Costs could increase with an expansion of the program as the supply of volunteer mentors who work for free is likely to be limited. In addition, paid mentors might be less prosocial than volunteer mentors, such that the benefits of the program would also decrease in such a case.

2000; Putnam, 2015), our results provide evidence that the role model aspect is key when it comes to children's development of prosociality and the distribution of life chances.

Finally, our study contributes to the debate on cultural evolution. Our results provide evidence in favor of theories of human social behaviors that do not exclusively rely on genetic causes but rather on patterns of social interaction such as teaching, imitation and learning in social environments (Boyd and Richerson, 1985; Fehr and Fischbacher, 2003; Henrich et al., 2004; Boyd, Richerson, and Henrich, 2011). Previous support for theories of cultural evolution primarily relies on cross society comparisons. We contribute complementary evidence by showing substantial within-society heterogeneities in prosociality arising from randomly-assigned social environments.

References

Aizer, Anna and Janet Currie. 2014. "The Intergenerational Transmission of Inequality: Maternal Disadvantage and Health at Birth." *Science* 344 (6186):856–861.

Alan, Sule, Nazli Baydar, Teodora Boneva, Thomas Crossley, and Seda Ertac. 2017. "Transmission of Risk Preferences from Mothers to Daughters." *Journal of Economic Behavior & Organization* 134 (1):60–77.

Alan, Sule and Seda Ertac. 2017. "Belief in Hard Work and Altruism: Evidence from a Randomized Experiment." Mimeo.

———. 2018. "Fostering Patience in the Classroom: Results from a Randomized Educational Intervention." *Journal of Political Economy,* forthcoming.

Algan, Yann, Elizabeth Beasley, Frank Vitaro, and Richard E. Tremblay. 2016. "The Impact of Non-Cognitive Skills Training on Academic and Non-academic Trajectories: From Childhood to Early Adulthood." Mimeo.

Almås, Ingvild, Alexander W. Cappelen, Erik Ø. Sørensen, and Bertil Tungodden. 2010. "Fairness and the Development of Inequality Acceptance." *Science* 328 (5982):1176–1178.

Attanasio, Orazio P., Camila Fernández, Emla O.A. Fitzsimons, Sally M. Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina. 2014. "Using the Infrastructure of a Conditional Cash Transfer Program to Deliver a Scalable Integrated Early Child Development Program in Colombia: Cluster Randomized Controlled Trial." *BMJ* 349 (g5785):1–12.

Bauer, Michal, Julie Chytilová, and Barbara Pertold-Gebicka. 2014. "Parental Background and Other-Regarding Preferences in Children." *Experimental Economics* 17 (1):24–46.

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse. 2012. "The Relationship Between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4 (1):453–478.

Becker, Gary S. 1976. "Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology." *Journal of Economic Literature* 14 (3):817–826.

Bénabou, Roland and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5):1652–1678.

Benenson, Joyce F., Joanna Pascoe, and Nicola Radmore. 2007. "Children's Altruistic Behavior in the Dictator Game." *Evolution and Human Behavior* 28 (3):168–175.

Bisin, Alberto and Thierry Verdier. 2000. "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits." *The Quarterly Journal of Economics* 115 (3):955–988.

———. 2001. "The Economics of Cultural Transmission and the Dynamics of Preferences." *Journal of Economic Theory* 97 (2):298–319.

Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas Ter Weel. 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43 (4):972–1059.

Boyd, Robert and Peter J. Richerson. 1985. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.

Boyd, Robert, Peter J. Richerson, and Joseph Henrich. 2011. "The Cultural Niche: Why Social Learning is Essential for Human Adaptation." *Proceedings of the National Academy of Sciences* 108 (Suppl 2):10918–10925.

Burks, Stephen V., Daniele Nosenzo, Jon Anderson, Matthew Bombyk, Derek Ganzhorn, Lorenz Götte, and Aldo Rustichini. 2016. "Lab Measures of Other-Regarding Preferences Can Predict Some Related on-the-job Behavior: Evidence from a Large Scale Field Experiment." IZA Discussion Paper No. 9767.

Campbell, Frances, Gabriella Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan. 2014. "Early Childhood Investments Substantially Boost Adult Health." *Science* 343 (6178):1478–1485.

Cappelen, Alexander W, John A List, Anya Samek, and Bertil Tungodden. 2016. "The Effect of Early Education on Social Preferences." NBER Working Paper No. 22898.

Carpenter, Jeffrey and Erika Seki. 2011. "Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyama Bay." *Economic Inquiry* 49 (2):612–630.

Carrell, Scott E, Marianne E Page, and James E West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *The Quarterly Journal of Economics* 125 (3):1101–1144.

Case, Anne, Darren Lubotsky, and Christina Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review* 92(5):1308–1334.

Cesarini, David, Christopher T. Dawes, James H. Fowler, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace. 2008. "Heritability of Cooperative Behavior in the Trust Game." *Proceedings of the National Academy of Sciences* 105 (10):3721–3726.

Cesarini, David, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace. 2009. "Genetic Variation in Preferences for Giving and Risk Taking." *The Quarterly Journal of Economics* 124 (2):809–842.

Charness, Gary and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *The Quarterly Journal of Economics* 117 (3):817–869.

Cooper, David and John H. Kagel. 2009. "Other Regarding Preferences: A Selective Survey of Experimental Results." *Handbook of Experimental Economics* .

Cunha, Flavio and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2):31–47.

Currie, Janet and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *The Quarterly Journal of Economics* 118 (4):1495–1532.

Deckers, Thomas, Armin Falk, Fabian Kosse, Pia Pinger, and Hannah Schildberg-Hörisch. 2017. "Socio-Economic Status and Inequalities in Children's IQ and Economic Preferences." IZA Discussion Paper No. 11158.

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market." *The Quarterly Journal of Economics* 132 (4):1593–1640.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2009. "Homo Reciprocans: Survey Evidence on Behavioural Outcomes." *The Economic Journal* 119 (536):592–612.

———. 2012. "The Intergenerational Transmission of Risk and Trust Attitudes." *The Review of Economic Studies* 79 (2):645–677.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences." *Journal of the European Economic Association* 9 (3):522–550.

Doyle, Orla, Colm Harmon, James J. Heckman, Caitriona Logue, and Seong Hyeok Moon. 2017. "Early Skill Formation and the Efficiency of Parental Investment: A Randomized Controlled Trial of Home Visiting." *Labour Economics* 45 (1):40–58.

Duncan, Greg J. and Katherine Magnuson. 2013. "Investing in Preschool Programs." *The Journal of Economic Perspectives* 27 (2):109–132.

Eby, Lillian T., Tammy D. Allen, Sarah C. Evans, Thomas Ng, and David L. DuBois. 2008. "Does Mentoring Matter? A Multidisciplinary Meta-analysis Comparing Mentored and Non-mentored Individuals." *Journal of Vocational Behavior* 72 (2):254–267.

Eisenberg, Nancy, Richard A. Fabes, and Tracy L. Spinrad. 2006. *Handbook of Child Psychology: Social, Emotional, and Personality Development*, chap. Prosocial development. 646–718.

Eisenberg, Nancy, Tracy L. Spinrad, and Ariel Knafo-Noam. 2015. *Handbook of Child Psychology and Developmental Science: Socioemotional Processes*, chap. Prosocial development. 610–656.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2016. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." IZA Discussion Paper No. 9674.

Falk, Armin, Anke Becker, Thomas J. Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *Quarterly Journal of Economics,* forthcoming.

Falk, Armin and Urs Fischbacher. 2006. "A Theory of Reciprocity." *Games and Economic Behavior* 54 (2):293–315.

Fehr, Ernst, Helen Bernhard, and Bettina Rockenbach. 2008. "Egalitarianism in Young Children." *Nature* 454 (7208):1079–1083.

Fehr, Ernst and Urs Fischbacher. 2003. "The Nature of Human Altruism." *Nature* 425 (6960):785–791.

Fehr, Ernst, Urs Fischbacher, Bernhard von Rosenbladt, Jürgen Schupp, and Gert G. Wagner. 2002. "A Nation-Wide Laboratory." *Journal of Applied Social Science Studies* 122:519–542.

Fehr, Ernst and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415 (6868):137–140.

Fehr, Ernst, Daniela Glätzle-Rützler, and Matthias Sutter. 2013. "The Development of Egalitarianism, Altruism, Spite and Parochialism in Childhood and Adolescence." *European Economic Review* 64 (1):369–383.

Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics* 114 (3):817–868.

———. 2006. "The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories." *Handbook of the Economics of Giving, Altruism and Reciprocity* .

Foster, Lisa. 2001. "Effectiveness of Mentor Programs: Review of the Literature from 1995 to 2000." Tech. rep., California State Library, Sacramento. California Research Bureau. Prepared at the Request of Senator Dede Alpert, Chair, Senate Select Committee On Family, Child and Youth Development.

Frölich, Markus and Martin Huber. 2017. "Direct and Indirect Treatment Effects–Causal Chains and Mediation Analysis with Instrumental Variables." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79 (5):1645––1666.

Ganzeboom, Harry B.G., Paul M. De Graaf, and Donald J. Treiman. 1992. "A Standard International Socio-economic Index of Occupational Status." *Social Science Research* 21 (1):1–56.

Gerlitz, Jean-Yves and Jürgen Schupp. 2005. "Zur Erhebung der Big-Five-basierten Persoenlichkeitsmerkmale im SOEP." *DIW Research Notes* 4.

Goodman, Robert. 1997. "The Strengths and Difficulties Questionnaire: A Research Note." *Journal of Child Psychology and Psychiatry* 38 (5):581–586.

Grossman, Jean Baldwin and Joseph P. Tierney. 1998. "Does Mentoring Work? An Impact Study of the Big Brothers Big Sisters Program." *Evaluation Review* 22 (3):403–426.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Cultural Biases in Economic Exchange?" *The Quarterly Journal of Economics* 124 (3):1095–1131.

Harbaugh, William T. and Kate Krause. 2000. "Children's Altruism in Public Good and Dictator Experiments." *Economic Inquiry* 38 (1):95–109.

Hawkins, J. David, Richard F. Catalano, Rick Kosterman, Robert Abbott, and Karl G. Hill. 1999. "Preventing Adolescent Health-Risk Behaviors by Strengthening Protection during Childhood." *Archives of Pediatrics & Adolescent Medicine* 153 (3):226–234.

Hawkins, J. David, Rick Kosterman, Richard F. Catalano, Karl G. Hill, and Robert D. Abbott. 2005. "Promoting Positive Adult Functioning through Social Development Intervention in Childhood: Long-Term Effects from the Seattle Social Development Project." *Archives of Pediatrics & Adolescent Medicine* 159 (1):25–31.

Heckman, James, Seong Hyeok Moon, Rodrigo Pinto, Peter Savelyev, and Adam Yavitz. 2010a. "Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program." *Quantitative Economics* 1 (1):1–46.

Heckman, James J. 2008. "Schools, Skills, and Synapses." *Economic Inquiry* 46 (3):289–324.

———. 2013. *Giving Kids a Fair Chance*. MIT Press, Cambridge.

Heckman, James J and Jorge Luís García. 2017. "Social Policy: Targeting Programmes Effectively." *Nature Human Behaviour* 1 (1):0019.

Heckman, James J, Seong Hyeok Moon, Rodrigo Pinto, Peter A Savelyev, and Adam Yavitz. 2010b. "The Rate of Return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94 (1-2):114–128.

Heckman, James J and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annual Review of Economics* 6 (1):689–733.

Heckman, James J., Rodrigo Pinto, and Peter A. Savelyev. 2013. "Understanding the Mechanisms through which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103 (6):2052–2086.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24 (3):411–482.

Heinz, Matthias and Heiner Schumacher. 2017. "Signaling Cooperation." *European Economic Review* 98:199–216.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis. 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-scale Societies.* Oxford University Press.

Hertwig, Ralph and Andreas Ortmann. 2001. "Experimental Practices in Economics: A Methodological Challenge for Psychologists?" *Behavioral and Brain Sciences* 24 (3):383–403.

John, Katrin and Stephan L Thomsen. 2015. "School-track Environment or Endowment: What Determines Different Other-regarding Behavior Across Peer Groups?" *Games and Economic Behavior* 94:122–141.

Kautz, Tim, James J. Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans. 2014. "Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success." OECD Education Working Paper No. 110.

Keele, Luke, Dustin Tingley, and Teppei Yamamoto. 2015. "Identifying Mechanisms Behind Policy Interventions via Causal Mediation Analysis." *Journal of Policy Analysis and Management* 34 (4):937–963.

Knack, Stephen and Philip Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *The Quarterly Journal of Economics* 112 (4):1251–1288.

Kofoed, Michael S. and Elizabeth McGovney. 2017. "The Effect of Same-Gender and Same-Race Role Models on Occupation Choice: Evidence from Randomly Assigned Mentors at West Point." *Journal of Human Resources,* forthcoming.

Kosfeld, Michael, Markus Heinrichs, Paul J. Zak, Urs Fischbacher, and Ernst Fehr. 2005. "Oxytocin Increases Trust in Humans." *Nature* 435 (7042):673–676.

Kosse, Fabian and Friedhelm Pfeiffer. 2012. "Impatience among Preschool Children and their Mothers." *Economics Letters* 115 (3):493–495.

La Porta, Rafael, Florencio Lopez-De-Silane, Andrei Shleifer, and Robert W. Vishny. 1997. "Trust in Large Organizations." *American Economic Review* 87 (2):333–338.

Lang, Frieder R., Dennis John, Oliver Lüdtke, Jürgen Schupp, and Gert G. Wagner. 2011. "Short Assessment of the Big Five: Robust across Survey Methods except Telephone Interviewing." *Behavior Research Methods* 43 (2):548–567.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies* 76 (3):1071–1102.

MacCallum, Judith and Susan Beltman. 2002. "Role Models for Young People: What Makes an Effective Role Model Program." A Report to the Australian Clearing House for Youth Studies.

McCord, Joan. 1978. "A Thirty-Year Follow-Up of Treatment Effects." *American Psychologist* 33 (3):284–289.

Moodie, Marjory L. and Jane Fisher. 2009. "Are Youth Mentoring Programs Good Value-For-Money? An Evaluation of the Big Brothers Big Sisters Melbourne Program." *BMC Public Health* 9 (1):41.

Müller-Kohlenberg, H. and S. Drexler. 2013. *Mentoring: Practices, Potential Challenges and Benefits*. Nova Science Publishers, New York.

Nelson, Charles A., Charles H. Zeanah, Nathan A. Fox, Peter J. Marshall, Anna T. Smyke, and Donald Guthrie. 2007. "Cognitive Recovery in Socially Deprived Young Children: The Bucharest Early Intervention Project." *Science* 318 (5858):1937–1940.

OECD. 2017. "OECD stat." Database.

Ostrom, Elinor, Thomas Dietz, Nives Dolsak, Paul C. Stern, Susan Stonich, and Elke U. Weber. 2002. *The Drama of the Commons*. National Academies Press, Washington D.C.

Putnam, Robert D. 2015. *Our Kids: The American Dream in Crisis*. Simon and Schuster, New York.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *The American Economic Review* :1281–1302.

Rao, Gautam. 2014. "Familiarity Does not Breed Contempt: Generosity, Discrimination and Diversity in Delhi Schools." Mimeo.

Rodríguez-Planas, Núria. 2012. "Mentoring, Educational Services, and Incentives to Learn: What Do We Know about Them?" *Evaluation and Program Planning* 35 (4):481–490.

Romano, Joseph P. and Michael Wolf. 2005. "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing." *Journal of the American Statistical Association* 100 (469):94–108.

———. 2016. "Efficient Computation of Adjusted p-values for Resampling-based Stepdown Multiple Testing." *Statistics & Probability Letters* 113 (1):38–40.

Sandner, Malte and Tanja Jungmann. 2016. "How Much Can We Trust Maternal Ratings of Early Child Development in Disadvantaged Samples?" *Economics Letters* 141 (1):73–76.

Skinner, Burrhus Frederic. 1953. *Science and Human Behavior*. Macmillan, New York.

SOEP. 2011. "German Socio-Economic Panel, Data for Years 1984-2009, Version 26." doi:10.5684/soep.v26. Dataset.

Stuart, Elizabeth A., Stephen R. Cole, Catherine P. Bradshaw, and Philip J. Leaf. 2011. "The Use of Propensity Scores to Assess the Generalizability of Results from Randomized Trials." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 174 (2):369–386.

Sutter, Matthias and Martin G. Kocher. 2007. "Trust and Trustworthiness across Different Age Groups." *Games and Economic Behavior* 59 (2):364–382.

Tierney, Joseph P. and Jean Baldwin Grossman. 2000. *Making a Difference - An Impact Study of Big Brothers Big Sisters*.

United States General Accounting Office. 1991. "Discount Rate Policy." Government Printing Office, Washington, DC.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp. 2007. "The German Socio-Economic Panel Study (SOEP)–Scope, Evolution and Enhancements." *Journal of Applied Social Science Studies* 127 (1):139–169.

Werner, Emmy E. 2000. "Protective Factors and Individual Resilience." *Handbook of Early Childhood Intervention* .

Williamson, Rebecca A, Meghan Rose Donohue, and Erin C Tully. 2013. "Learning How to Help Others: Two-year-olds' Social Learning of a Prosocial Act." *Journal of Experimental Child Psychology* 114 (4):543–550.

Yamamoto, Teppei. 2013. "Identification and Estimation of Causal Mediation Effects with Treatment Noncompliance." Mimeo.

Zumbuehl, Maria, Thomas Dohmen, and Gerard Pfann. 2013. "Parental Investment and the Intergenerational Transmission of Economic Preferences and Attitudes." IZA Discussion Paper No. 7476.

Appendix

A Additional Tables and Figures

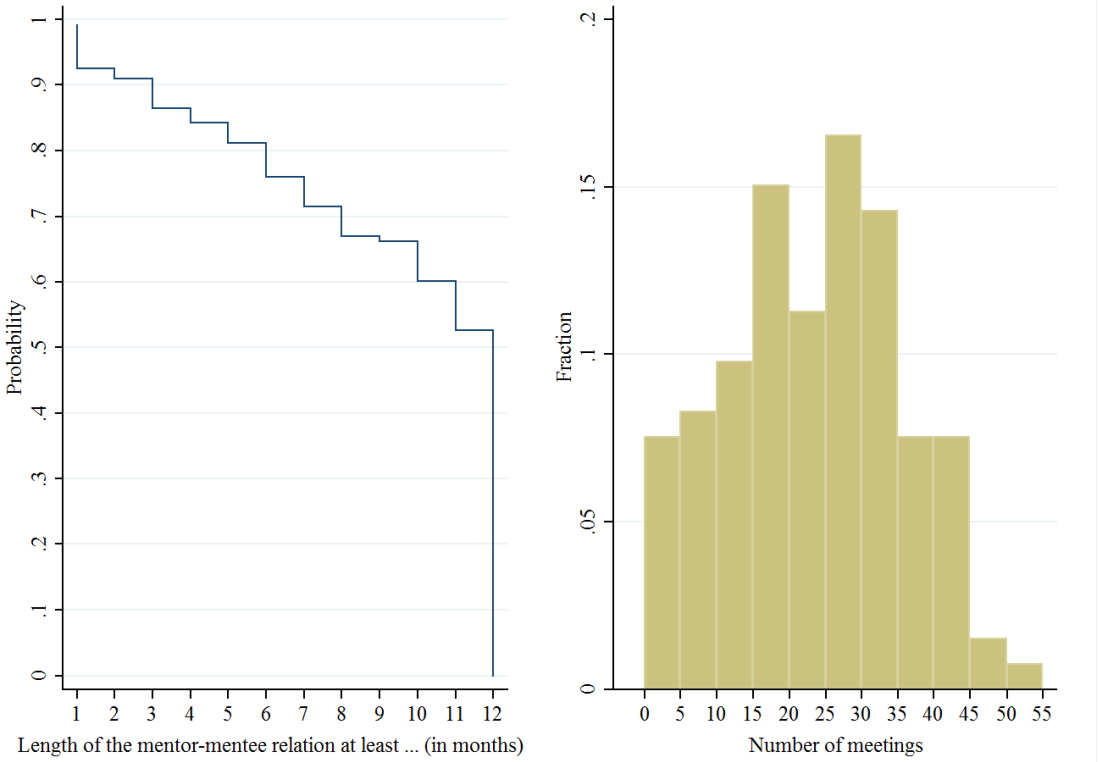

Figure A1: Length of mentor-mentee relations and numbers of meetings. The left panel shows a reverse cumulative probability plot. It indicates that more than 50% of the pairs met for twelve months, while more than 80% of the pairs met at least for six months. The mean is 9.3 months. The right panel shows a histogram of the number of meetings. The number of meetings for most of the pairs is centered around the mean of 22.8.

Figure A2: Toys arranged in four categories (example).

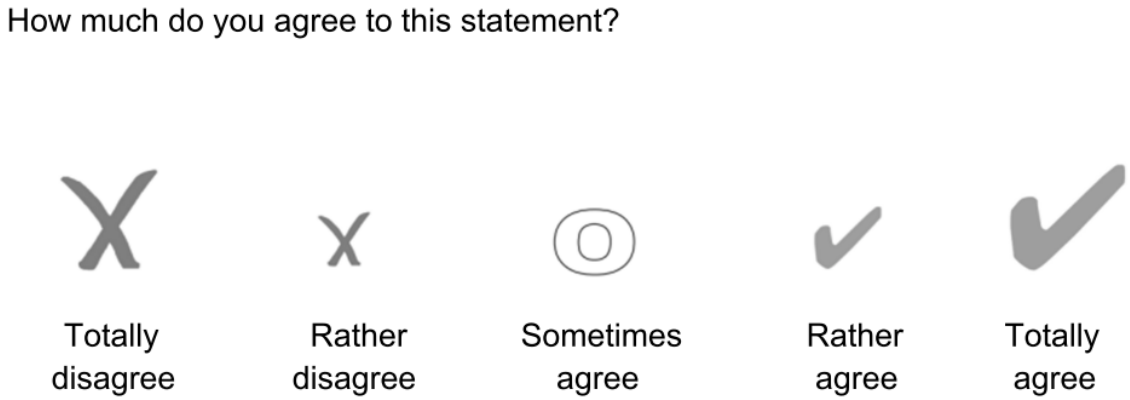

Figure A3: Rating scale used in the children survey.

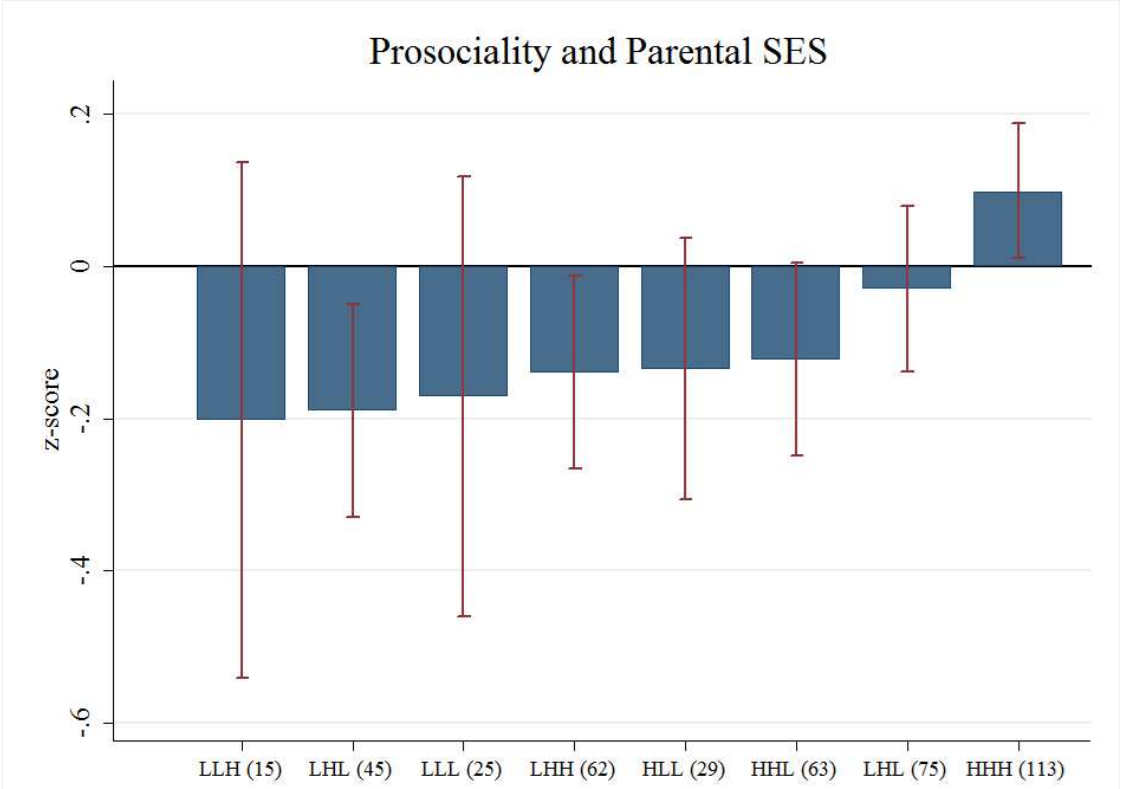

Figure A4: Sub-groups of combinations of parental SES characteristics (ordered by means of children's prosociality). The number of observations for every sub-group is displayed in brackets (in sum $N = 427$, Control High SES and Control Low SES). The prosociality measure of children is constructed as equally-weighted scores of standardized measures of altruism, trust and other-regarding behavior, respectively. The scale on the y-axis indicates z-scores (i.e., standardized measures) of children's prosociality. Error bars show standard errors of the means (SEM).

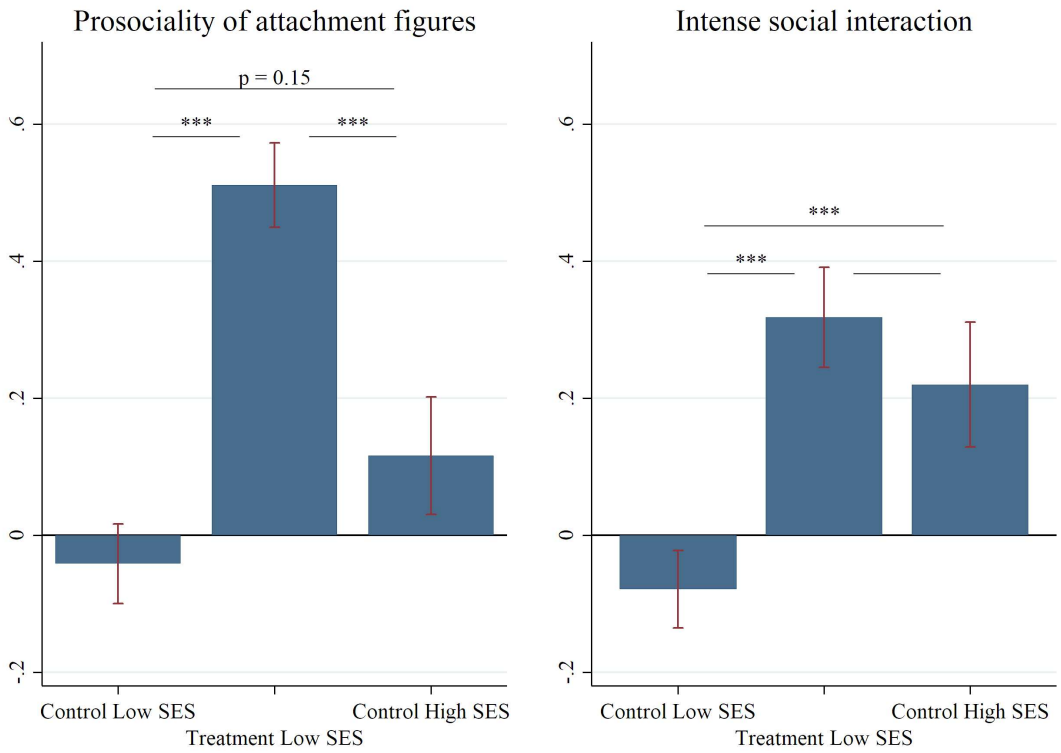

Figure A5: This graph compares investments for children in all three groups: Control Low SES, Treatment Low SES and Control High SES. The measures are compared in terms of z-scores (standardized on distribution of Control Low SES and Control High SES). For details on the measures, see section 2.3. For children in the treatment group the measures aggregate investments received at home and during the intervention, for details see 3.3. Error bars show standard errors of the means. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively (two-sided t-tests).

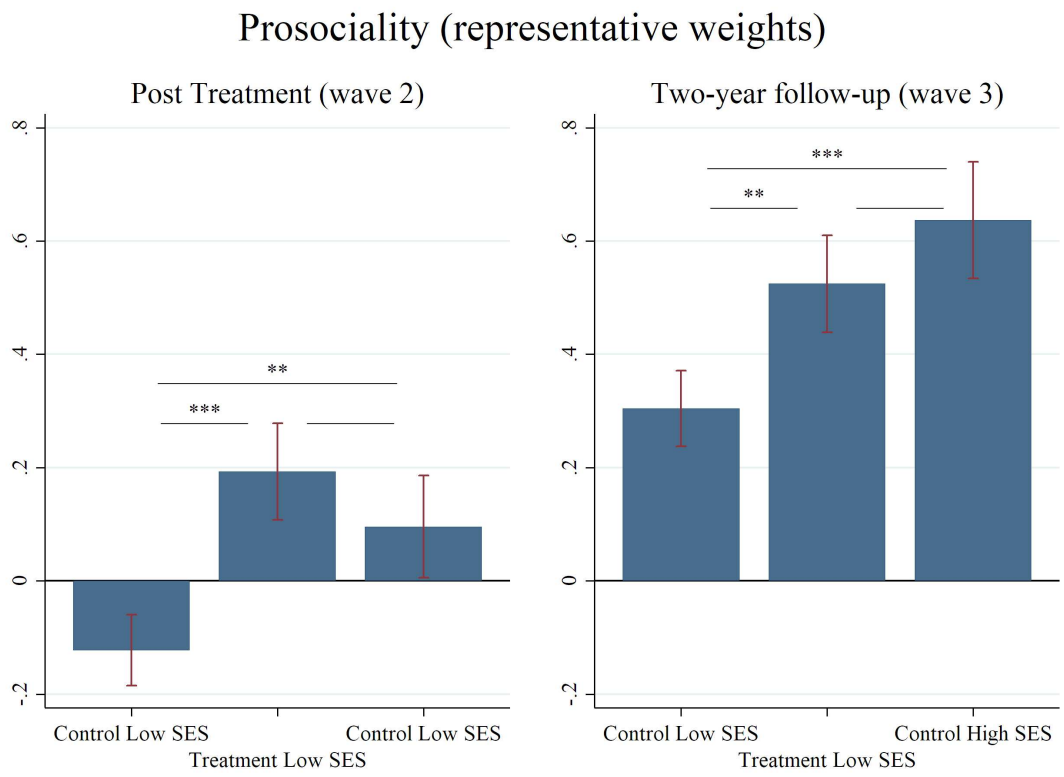

Figure A6: Main analysis using representative weights: Significantly higher levels of prosociality for treated children compared to untreated children (Treatment Low SES vs. Control Low SES). There is no significant difference between Treatment Low SES and Control High SES. The scale on the y-axis indicates z-scores (i.e., standardized measures) of children's prosociality. Standardization is conducted using the distribution of wave 2. Error bars show White robust standard errors. Data is re-weighted according to a representative sample of families in Germany. Weights take maternal prosociality, the intensity of mother-child interactions and different dimensions of SES into account. For details on the construction of the weights, see section B.4.1. The prosociality measure is constructed as equally-weighted score of standardized measures of altruism, trust and other-regarding behavior. *** and ** indicate significant differences at the 1 and 5 percent level, respectively (two-sided t-tests).

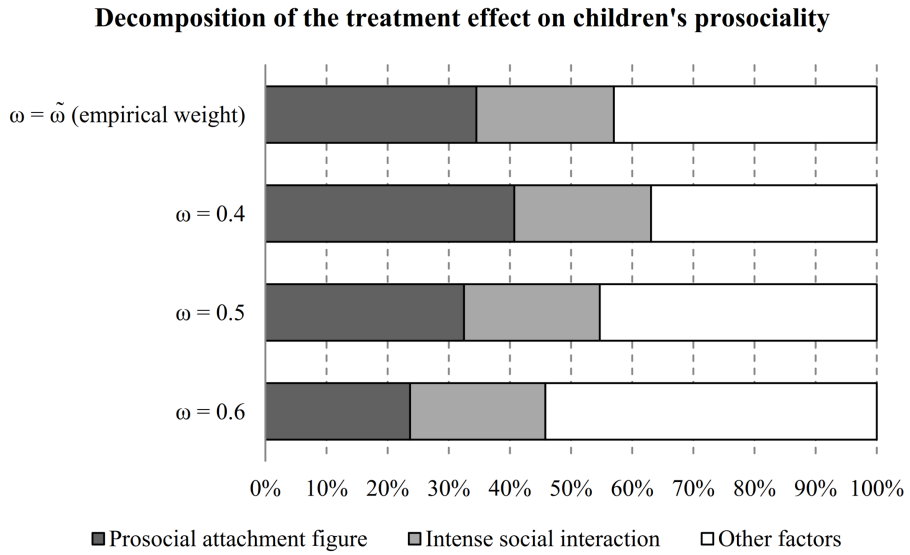

Figure A7: The figure shows the estimates of the decomposition as described in equation 2. Each bar represents the total treatment effect normalized to 100%. The dark gray area indicates the share of the treatment effect explained by changes in the prosocial attachment figure. The light gray area indicates the share of the treatment effect explained by the experience of intense social interaction. The white area indicates unmeasured other factors. The bar in line 1 (empirical weights, see section B.5 for details) is the same as shown in figure 3. In lines 2 to 4, we repeat the analysis using ad-hoc weights instead of the empirical values of ω. For equal weights ($\omega = 0.5$), our analysis yields similar results. For $\omega = 0.4$ ($\omega = 0.6$), the importance of the prosocial attachment figure slightly increases (decreases). Two-sided p-values: $p < 0.01$ for all channels in all four decompositions. For details on the estimation, see Heckman, Pinto, and Savelyev (2013).

	Prosociality	Altruism	Trust	Other-reg. behavior	PS w/o o-r beh.
	(1)	(2)	(3)	(4)	(5)
Treatment Dummy	0.294***	0.199**	0.235**	0.108	0.305***
	[0.093]	[0.095]	[0.098]	[0.097]	[0.096]
Constant	-0.191*	-0.147	-0.167*	-0.039	-0.220**
	[0.100]	[0.104]	[0.101]	[0.098]	[0.104]
Observations	494	493	494	492	494
R^2	0.018	0.008	0.012	0.003	0.020

Table A1: Treatment effects for the joint measure and facets of prosociality conditional on city of residence. Coefficients are OLS estimates based on post-treatment data (wave 2), with White robust standard errors in brackets. Regressions include a dummy variable (living in Cologne equals 1) to account for the fact that the assignment into treatment was random conditional on city of residence. The prosociality measure in column (1) is constructed as equally-weighted score of standardized measures of altruism, trust and other-regarding behavior. The prosociality measure in column (5) is constructed as equally-weighted score of standardized measures of altruism and trust only (without other-regarding behavior). ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively. A joint F-test rejects the null-hypothesis of a zero treatment effect for the three facets of prosociality (altruism, trust, other-regarding behavior; $p = 0.010, F = 3.78, N = 491$).

	Prosociality	
	(1)	(2)
Treatment Dummy	0.300***	0.250***
	[0.093]	[0.092]
Prosociality of mother	0.245***	
	[0.056]	
Prosociality of mother x Treat	-0.214**	
	[0.091]	
Social interaction		0.255***
		[0.057]
Social interaction x Treat		-0.192**
		[0.094]
Constant	-0.134**	-0.100*
	[0.057]	[0.057]
Observations	484	491
R^2	0.060	0.057

Table A2: Interaction effects of treatment and parental background. The prosociality measures of mother and child are constructed as equally-weighted scores of standardized measures of altruism, trust and other-regarding behavior, respectively. The intensity of social interaction of mother and child is the average share of highly-interactive activities per afternoon. Prosociality and social interaction measures are standardized. The significant negative interaction terms in columns (1) and (2), "Prosociality of mother x Treat" and "Social interaction x Treat", indicate that the treatment benefits those children most who experience little social interaction and whose mothers score low on prosociality. Displayed coefficients are OLS estimates with White robust standard errors in brackets. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	Prosociality			
	Wave 2		Wave 3	
	(1)	(2)	(3)	(4)
Treatment Dummy	0.273***	0.315***	0.217**	0.220**
	(0.093)	(0.106)	(0.103)	(0.109)
Constant	-0.122**	-0.122*	0.338***	0.304***
	(0.058)	(0.063)	(0.060)	(0.067)
Weights	No	Representative	No	Representative
Observations	494	494	411	411
R^2	0.017	0.023	0.011	0.012

Table A3: Treatment effects for prosociality in waves 2 and 3 using inverse probability weighting (weights: representative population). Coefficients are weighted least-square (IPW) estimates with White robust standard errors in brackets. The dependent variable in columns (1) and (2) is the prosociality measure collected in wave 2. The dependent variable in columns (3) and (4) is the prosociality measure collected in wave 3 (two-year follow-up). The data are re-weighted according to a representative sample of families in Germany. Weights take maternal prosociality, intensity of mother-child interaction and different dimensions of SES into account. For details on the construction of the weights, see section B.4.1. The prosociality measure is constructed as equally-weighted score of standardized measures of altruism, trust and other-regarding behavior. Dependent variables are standardized. Standardization is conducted using the distribution of wave 2. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	Assigned to treatment	Lost to follow-up	
		Wave 2	Wave 3
	(1)	(2)	(3)
Prosociality (wave 1)	0.009	0.015	-0.007
	[0.020]	[0.020]	[0.024]
Treatment Dummy		-0.022	-0.003
		[0.032]	[0.040]
PS x Treatment Dummy		-0.021	-0.026
		[0.031]	[0.039]
Constant	0.361***	0.172***	0.300***
	[0.020]	[0.020]	[0.024]
Restricted to low SES	Yes	Yes	Yes
Observations	590	590	590
R^2	0.000	0.002	0.002

Table A4: Checks for baseline balance and selective attrition. In column (1), we test for baseline balance. The dependent variable is one if a child was selected into the Treatment Low SES group and zero if selected into the Control Low SES group. In columns (2) and (3), we test for selective attrition. The dependent variable is one if a child is lost to follow-up, i.e., did not take part in the wave 2 or 3 interviews, and zero otherwise. The explanatory variable prosociality is constructed as equally-weighted score of standardized measures of altruism, trust and other-regarding behavior. Standardization is conducted using the distribution of wave 2. All measures in wave 1 were collected before the treatment assignment took place. Displayed coefficients are OLS estimates with White robust standard errors in brackets. ***, **, * indicate significance at 1, 5 and 10 percent level, respectively.

	Prosociality	
	Wave 2	Wave 3
	(1)	(2)
Treatment Dummy	0.273***	0.214**
	[0.092]	[0.104]
Constant	-0.118**	0.332***
	[0.058]	[0.060]
Weights	Baseline Sample	Baseline Sample
Observations	494	411
R^2	0.017	0.011

Table A5: Treatment effects for prosociality in waves 2 and 3 using inverse probability weighting (weights: baseline sample). Coefficients are weighted least-square (IPW) estimates with White robust standard errors in brackets. The dependent variable in column (1) is the prosociality measure collected in wave 2. The dependent variable in column (2) is the prosociality measure collected in wave 3 (two-year follow-up). Weights are predicted inverse probabilities of not being lost to follow-up. For details on the construction of the weights, see section B.4.2. The prosociality measure is constructed as equally-weighted score of standardized measures of altruism, trust and other-regarding behavior. Dependent variables are standardized. Standardization is conducted using the distribution of wave 2. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	Prosociality			
	Wave 2 bounds		Wave 3 bounds	
	Lower	Upper	Lower	Upper
	(1)	(2)	(3)	(4)
Treatment Dummy	0.225**	0.326***	0.207*	0.217*
	[0.104]	[0.108]	[0.122]	[0.124]
Number of obs.	590	590	590	590
Number of selected obs.	494	494	411	411

Table A6: Treatment effect bounds for prosociality in waves 2 and 3. The bounds are estimated using the trimming procedure suggested by Lee (2009). Standard errors (in parenthesis) are bootstrapped using 1,000 bootstrap replications. The prosociality measure is constructed as equally-weighted score of standardized measures of altruism, trust and other-regarding behavior. The dependent variable in columns (1) and (2) is standardized post-treatment prosociality (wave 2). The dependent variable in columns (3) and (4) is standardized two-year follow-up prosociality (wave 3). Standardization is conducted using the distribution of wave 2. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	Prosociality	
	(1)	(2)
Treatment Dummy	0.254***	0.202**
	[0.095]	[0.091]
Dummy Wave 2	0.092	0.111**
	[0.057]	[0.056]
Dummy Wave 3		0.467***
		[0.065]
Constant	-0.179***	-0.175***
	[0.021]	[0.026]
Individual Fixed Effects	Yes	Yes
Considered waves	1 and 2	1, 2 and 3
Observations	1,084	1,495

Table A7: Difference-in-differences treatment effects. Columns (1) and (2) show the coefficients of individual fixed-effects regressions with clustered standard errors (at the individual level) in brackets. The treatment dummy is zero for control group children in all waves, as well as for ITT children in wave 1. It is one for ITT children in waves 2 and 3. The dependent variable (prosociality) is constructed as equally-weighted score of standardized measures of altruism, trust and other-regarding behavior, respectively. It is standardized using the distribution of wave 2. In order to account for the fact that the assignment into treatment was random conditional on city of residence, time-varying city effects (base: Cologne) are differenced out from the dependent variable, i.e., wave-specific city fixed effects are estimated and subtracted. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

	Prosociality	Altruism	Trust	Other-regard. behavior
	(1)	(2)	(3)	(4)
LATE	0.398***	0.270**	0.318**	0.146
	[0.126]	[0.129]	[0.132]	[0.131]
Constant	-0.190*	-0.147	-0.166*	-0.038
	[0.100]	[0.104]	[0.101]	[0.098]
Observations	494	493	494	492

Table A8: Local average treatment effect (LATE) analysis using random group assignment as instrument for actual treatment. 133 of the 180 children who we intended to treat were actually matched with a mentor. The prosociality measure is constructed as equally-weighted score of standardized measures of altruism, trust and other-regarding behavior. All dependent variables are standardized. Coefficients are two-stage least squares (2SLS) estimates using random assignment as an instrument for actual treatment, White robust standard errors in brackets. Regressions include a dummy variable (living in Cologne equals 1) to account for the fact that the assignment into treatment was random conditional on city of residence. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively. A joint F-test rejects the null-hypothesis of a zero treatment effect for the three components of prosociality (altruism, trust, other-regarding behavior; $p = 0.010, F = 3.78, N = 491$).

Mentor-mentee characteristics	Standardized quality of mentor-mentee relation (1)	(2)	(3)	Δ child prosociality PS(wave 2) - PS(wave 1) (4)	(5)	(6)
Sex composition						
Same sex	0.339*			0.186		
	[0.199]			[0.280]		
Neighborhood						
Distance between places of residence		-0.038			-0.228**	
		[0.080]			[0.092]	
Parental education of mentor - mentee						
Base: high - low						
Low - high			0.307			-0.220
			[0.369]			[0.280]
Low - low			-0.054			0.222
			[0.488]			[0.523]
High - high			0.004			-0.181
			[0.278]			[0.230]
Observations	133	130	86	133	130	86
R^2	0.020	0.007	0.039	0.015	0.066	0.042

Table A9: The dependent variable in columns (1) to (3) is the standardized quality of the mentor-mentee relation as indicated by the mentor, see also section B.5. The dependent variable in columns (4) to (6) is the difference in standardized prosociality between wave 2 and wave 1. Coefficients are OLS estimates with White robust standard errors in brackets. Distance between places of residence is the standardized commuting time between the home of the mentor and the home of the mentee. In the analysis of parental education, the left out category is "mentor high - mentee low". Mentors and mentees are categorized as having a high parental background if at least one parent holds a university entry diploma. The lower number of observations are due to missing information on the mentors. All regressions also include a constant, city and gender dummies. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

B Additional data descriptions

B.1 SOEP data: Predicting life outcomes with prosociality data

In order to explore the predictive power of prosociality on different life outcomes, we use data from the German Socio-Economic Panel Study (SOEP), a large panel data set that is representative of the adult population living in Germany (see Wagner, Frick, and Schupp (2007) for a detailed description of the SOEP). We use information from seven waves collected in the years between 2003 and 2009 (SOEP, 2011). In each of these waves, more than 20,000 individuals were interviewed. The SOEP combines extensive socio-demographic information with various measures of attitudes, preferences and psychological traits. In particular, the SOEP includes survey items relating to the three facets of prosociality (altruism, trust, other-regarding behavior), which we focus on in the main part of this paper.

For the following analysis, we build on the definitions and variables used in Becker et al. (2012). The survey measure for *altruism* is the answer to the question of how important it is for the participant "to be there for others". Answers were given on a four-point scale. The altruism question was included in the 2004 and 2008 waves. Measures on *trust* were collected in the two waves 2003 and 2008, using three sub-statements about whether "one can trust people", whether "in these times one can't rely on anybody else" and whether "when dealing with strangers it is better to be cautious". Answers were given on a four-point scale ranging from "Totally agree" to "Totally disagree"(see, Fehr et al., 2002). As a proxy for *other-regarding behavior*, we use a measure of the Big Five dimension Agreeableness. For this purpose, we use a short questionnaire version of the Big Five, the BFI-S, which has been developed by Gerlitz and Schupp (2005) and was used in the 2005 and 2009 waves of the SOEP. The outcome data that we use stem from 2009. For further details, see Becker et al. (2012).

In order to reduce measurement error, we aggregate the measures of the facets over different waves in the SOEP if possible. As for the children (for a description, see section 2.3), the joint measure of prosociality is the equally-weighted score of the three standardized facets. For results on the three facets individually, see table 9 in Becker et al. (2012).

	Subj. Health (5-point Likert) (1)	**Happiness** (11-p. Likert) (2)	**Gross Wage** (hourly, in €) (3)	**Unempl.** (dummy) (4)	**Education** (in years) (5)
Std. Prosociality	0.149***	0.462***	0.847***	-0.022***	0.492***
	[0.007]	[0.014]	[0.120]	[0.003]	[0.022]
Age [in years]	-0.021***	-0.004***	0.254***	-0.001**	-0.017***
	[0.000]	[0.001]	[0.011]	[0.000]	[0.001]
Female dummy	-0.115***	-0.150***	-4.569***	0.022***	-0.543***
	[0.013]	[0.026]	[0.239]	[0.006]	[0.042]
Observations	17,741	17,737	8,807	11,241	16,967
R^2	0.181	0.065	0.108	0.006	0.047

Table B1: SOEP data from 2003 to 2009, exactly as in Becker et al. (2012). Coefficients are OLS estimates with White robust standard errors in brackets. All regressions also include a constant. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

B.2 Altruistic activities of mothers and mentors

In order to obtain information about revealed altruistic behavior for the mothers, we asked them in a follow-up survey in 2015 to indicate how often they engage in altruistic and non-altruistic activities in their leisure time. Altruistic activities comprise, e.g., helping a friend or charitable work. Non-altruistic activities include, e.g., visiting cultural events or exercising. The share of altruistic free time activities is 22.5%.

In order to obtain comparable information for the mentors, we asked them about their motives to take part in the program and rated motives as altruistic (pure benefit of the child) or non-altruistic (e.g., development of own personality or obtaining a certificate of participation). The share of altruistic motives is 70.8%, reflecting the recruiting strategy of the program.

Our joint measure of altruism of adults is the equally-weighted score of the standardized share of altruistic leisure time activities and the standardized rating on the question "How would you assess your willingness to share with others without expecting anything in return, for example your willingness to give to charity?" (Falk et al., 2016).

B.3 High-low SES gap in adults

We use representative SOEP data to investigate whether the SES gap in prosociality during childhood also exists for adults. In addition to the information on prosociality described in section B.1, the SOEP contains a measure of adults' *parental* SES, namely the well-established "International Socio-Economic Index of Occupational Status (ISEI)" developed by Ganzeboom, De Graaf, and Treiman (1992). The ISEI is based on a categorization of the International Standard Classification of Occupations (ISCO) codes, which are available for the parents of most SOEP participants. In order to obtain a binary SES classification comparable to the one in our study, we first aggregate maternal and paternal ISEI to a single parental ISEI.[39] We then conduct a median split and categorize an individual as having "low SES" if the parental ISEI falls in the bottom 50 percentile of his/her ten year birth cohort and as "high SES" otherwise.

The corresponding results are displayed in table B2. Adults with low SES parents score about 18% of a standard deviation lower than those with high SES parents. In terms of magnitude, these estimates closely align with those for elementary school children in section 3.1, e.g. 22.6% of a standard deviation in wave 2.

	Prosociality (of adults)	
	(1)	(2)
Low SES	-0.184***	-0.182***
	[0.018]	[0.018]
Age		-0.002***
		[0.001]
Female dummy		0.376***
		[0.018]
Observations	11,544	11,544
R^2	0.009	0.046

Table B2: SES gap in adulthood. The table displays coefficients of OLS regressions, in which the standardized measure of prosociality is regressed on a low SES dummy that equals 1 for low and 0 for high SES parents of adults. All regressions also include a constant. For details on the prosociality measure of adults, see section B.1. White robust standard errors are displayed in brackets. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

[39] We average the two ISEI if they are available for both parents and use the single available ISEI otherwise.

B.4 Robustness checks: weighting

B.4.1 Weighting scheme 1: representative population

In order to assess the representativeness of our results, we compare our sample to the population of families in Germany. The comparison comprises various dimensions of parental SES, as well as maternal prosociality and intensity of mother-child interaction. As a reference, we make use of data from the German Socio-Economic Panel (SOEP) (Wagner, Frick, and Schupp, 2007). The data are collected yearly and are representative of the German population. At present, the SOEP comprises more than 20,000 individuals in more than 10,000 households. We compare the families in our sample with children in the same age range in the SOEP. In order to compare the samples along the SES dimensions that we used in the sampling process, we elicited SES in the same way as in the SOEP (for details, see section 2.1). Our data structure further allows us to compare the families in the SOEP and in our panel in the two dimension that are key in the formation of prosociality of children: the prosociality of their mothers[40] and the intensity of mother-child interaction[41].
Table B3 shows how these dimensions compare between our sample and the SOEP. While the shares of low-income and single-parent families in our sample is higher than in the SOEP, the share of low-educated parents is lower. In order to compare the samples in terms of maternal prosociality and intensity of mother-child interaction, we standardized the measure using the SOEP as a reference, i.e., the mean in the SOEP is zero. For our sample, we find that mothers are more prosocial but interact less intensively with their child than mothers of children in the same age-range in the SOEP.
In order to check the extent to which these deviations from a representative sample affect our results, we use this information to re-weight our data based on propensity scores that model the probability of being in our sample. For this purpose, we estimate propensity scores based on a linear model including the SES dimensions, prosociality of the mother and intensity of mother-child interaction. The re-weighted

[40]Regarding agreeableness (Lang et al., 2011) and general trust (Fehr et al., 2002) the SOEP includes the same items that we use to construct maternal prosociality for the main analysis in this study (see also section 2.3). The SOEP does not include the same measures of altruism that we use in our data collection. Hence, to compare prosociality of mothers in the SOEP and in our sample, we construct a maternal prosociality proxy as an equally-weighted score of standardized agreeableness and standardized trust in both mother samples.

[41]The SOEP does not include time-spending data for the considered age range. Hence, we compare the samples regarding intensity of mother-child interaction using a proxy that is included in both surveys. Intensity of the mother-child interaction is proxied by a frequency rating of the mother regarding the statement "I show my child with words and gestures that I care about him. Mothers in both samples rated the statement on a 5-point Likert scale.

main results are displayed in table A3. The resulting effect sizes are very similar to those from the unweighted regression, suggesting that selective participation does not bias our results.

Variable	Families in the SOEP ($N = 739$)	Analyzed main sample ($N = 607$)
Share low HH equ income	30.0%	40.7%
Share low parental education	50.2%	38.2%
Share single parents	11.6%	37.1%
Std. maternal prosociality	0.000	0.300
Std. intensity of mother-child interaction	0.000	-0.295

Table B3: Comparison to a representative sample. For descriptions of sample and variables see section B.4.1. Within the sample of high SES families it also holds that the families in our sample are characterized by a higher degree of maternal prosociality (0.782 vs. 0.135) and by a lower intensity of mother-child interactions (-0.222 vs. -0.003).

B.4.2 Weighting scheme 2: attrition

After the one-year treatment period, all families who had participated in wave 1 were invited to take part in the post-treatment wave (wave 2). 85.3% of them (607 out of 712) took part in this second wave of interviews. This is our core sample. Two years after the end of the treatment period, all families who had participated in the second wave were invited to take part in the third wave of interviews. 83.9% (509 out of 607) took part in this two-year follow-up, see also figure 1. In order to check for selective attrition related to treatment status and children's prosociality (at baseline), we re-weight the observed data using inverse probability weighting (IPW). Weights are estimated from a Probit model of a binary selection indicator (indicating whether the wave 2 or 3 interview was conducted) regressed on baseline prosociality and treatment assignment.

The re-weighted main results are displayed in table A5. The effects sizes are very similar to the unweighted effect sizes, suggesting that selective attrition does not bias our results.

B.5 Program mediators

1. Highly-interactive activities: Given that we have detailed records of the mentor-child meetings and mother-child activities, it is an accounting exercise to measure time spending. Participation in the mentoring programm increases the share of highly-interactive interactions per afternoon (as a 12-month average) from 41.1% to 46.7% ($p < 0.01$, ITT effect).

2. (Empirical) weights of the attachment figures: The literature on role models (e.g., MacCallum and Beltman, 2002) suggests that the influence of role models is not necessarily driven by pure exposure, but rather by the "quality of the relationship" between child and role model. In the mediation analysis, we use the following relative weights:

Weight of the mother: (Quality of mother-child relation) / (Quality of mother-child relation + Quality of mentor-child relation).

Weight of the mentor: (Quality of mentor-child relation) / (Quality of mother-child relation + Quality of mentor-child relation)

Information on the *quality of the mentor-child relation* comes from the survey question answered by mentor: "How do you rate the quality of the Baloo-Mowgli relation?".[42] Information on the *quality of mother/parents-child relation* comes from the survey question answered by mother: "My child has a good relation to us parents." The relative weight of the mentor ranges from 28.6% to 80.0% (mean: 51.6%). For details on the prosocility measure of mothers and mentors, see section 2.3. For a comparison of the level of inputs of the different groups of children, see figure A5.

In order to conduct the decomposition on the full sample of children, missing values for mothers and mentors regarding activities and prosociality were imputed. There are few missing values for mothers ($< 2\%$). For those mothers, we simply use the group mean (e.g., mean of low SES mothers) to impute. For the mentors, we lack part of the information for about 40 individuals. In these cases, we predict the missing values using a fully-interacted model of mentor's age, mentor's gender and a rating of the mentor's supervisor about the effort of the mentor ("How do you rate the effort of the mentor?").

[42] Missings are imputed using predictions from a regression model where the self-rating is regressed on the quality rating of the supervisor ("How do you rate the quality of the Baloo-Mowgli relation?").

B.6 Comparison of mentors and mothers

We compare mentors and low SES mothers in the domains of prosociality, patience and willingness to take risk. We collected the same items from both groups. For a description of the prosocility measure, see section 2.3. Concerning time and risk preferences, we use questionnaire measures validated by Falk et al. (2016). For time preference, we use the measure: "When it comes to financial decisions, how do you assess your willingness to abstain from things today so that you will be able to afford more tomorrow. Please indicate on the scale, where the value 0 means 'not at all willing to abstain today' and the value 10 means 'very willing to abstain today' ". For risk preferences, we use the measure "How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please indicate on the scale, where the value 0 means: 'not at all willing to take risks' and the value 10 means: 'very willing to take risks' ". For a detailed discussed of this item, see also Dohmen et al. (2011).

Table B4 shows that low SES mothers do not differ from mentors in patience and willingness to take risk, while there is an enormous difference in prosociality by more than one standard deviation. This pattern is in line with what we expected given the self-selection of mentors into the program. Volunteering for the benefit of a child is a prosocial decision and not directly related to time and risk preferences.

Variable (z-scores)	Mentors	Low SES mothers	Sign. difference
Prosociality	1.655	-0.042	***
Patience	-0.172	-0.070	
Willingness to take risk	0.152	-0.006	

Table B4: Comparison of mentors and low SES mothers. This table compares mentors and low SES mothers. The measures are standardized. Standardizations are conducted using the wave 2 distribution of all mothers. For details on the prosociality measure, see section 2.3. For details on patience and willingness to take risk, see section B.6. The third column indicates results from two-sample two-sided t-tests. The numbers of observations are between 572 and 576 due to missing observations. ***, **, * indicate significance at the 1, 5 and 10 percent level, respectively.

B.7 Children's time and risk preferences

In order to assess whether the program also affected children's patience and willingness to take risk, we have collected incentivized measures of these dimensions. For a detailed description, see Deckers et al. (2017).

Patience of the children was measured using the "piggy bank" experiment. Children were endowed with seven 20 cent coins. They could choose how many coins to put in a piggy bank and how many to take out immediately. The amount put in the piggy bank was doubled and sent to the children via postal mail one week after the interview. The number of coins put into the piggy bank is our measure of the child's patience, where a higher number implies a higher degree of patience.

Willingness to take risk of the children was measured using the "coin-flipping" experiment. The children could decide between a safe and a risky option in two situations. The interviewer presented two coins in each of the two situations. In situation A, one of the coins had three stars printed on each side. The other coin had seven stars on one side and zero on the other. The interviewer explained that choosing the coin with three stars on each side implied winning three stars for certain. However, choosing the other coin implied that the outcome (seven or zero stars) was determined by chance, with both outcomes being equally likely. The safe amount (three stars) was also "determined" by a coin toss to reduce the likelihood that children chose the risky option only for entertainment or game value. Children chose which coin should be tossed. After children had made their decision, but before actually tossing the chosen coin, the interviewer presented two more coins in another color (situation B). Now, one coin had four stars on each side, while the other coin again had zero stars on one side and seven on the other. Children made their second decision and the interviewer tossed the two chosen coins. The order in which the two variations of the game (situation A versus situation B) were played was randomized. Our measure of willingness to take risk is the number of risky choices.

In table B5, columns (2) and (3), we explore the effect of the treatment on children's patience and willingness to take risk. For comparison, column (1) shows the effect on prosociality (as presented in table A1). Original p-values are shown in (parenthesis), Romano Wolf p-values are shown in square brackets, for details on the stepdown adjusted p-values robust to multiple hypothesis testing, see Romano and Wolf (2016)). Irrespective of which p-values we consider, in line with our role model hypothesis there is a highly significant treatment effect on prosociality but no significant effect on patience and willingness to take risk, for a discussion see 3.4.

	Prosociality	Patience	Willingness to take risk
	(1)	(2)	(3)
Treatment Dummy	0.294	0.105	0.000
Original p-values	(0.003)	(0.242)	(0.998)
Romano Wolf p-values	[0.007]	[0.432]	[0.997]
Observations	494	489	494

Table B5: Treatment effects on prosociality, patience and willingness to take risk. Coefficients are OLS estimates based on post-treatment data (wave 2). Original p-values are shown in (parenthesis), Romano Wolf p-values are shown in [square brackets], for details on the stepdown adjusted p-values robust to multiple hypothesis testing, see Romano and Wolf (2016). In the respective regressions, the outcome variable is regressed on a treatment dummy, a Cologne dummy (equaling zero for individuals living in Bonn) and a constant. For details on the prosociality measure, see section 2.3. For details on patience and willingness to take risk, see section B.7.

B.8 Benefit-cost ratios and the internal rate of return

Any reasonable approximation of the benefit-cost ratios of a program requires putting a dollar value on its gains. In section 3.4, we compute the lifetime labor market returns from participating in the mentoring program as a discounted stream of additional gross hourly earnings per year. Since our children have not yet entered the labor market, we use the prosociality estimates presented in table B1, which are based on a representative sample of German adults (for details, see Becker et al., 2012). This estimate, albeit not necessarily causal, is in line with the findings reported in Deming (2017). Furthermore, we assume that children on average enter the labor market when they are 23 years old and that they exit at the age of 66. Last, we assume that the returns are constant over the life cycle and that prosociality is equally productive among treated and control group children.

Given that program costs accrue immediately, while benefits accrue in the future, the resulting benefit-cost ratio depends on the discount rate used to ascertain the present value of future benefits. For example, with a treatment effect of 0.2 standard deviations, a (real) discount rate of 1% yields a benefit-cost ratio of 6.91, a discount rate of 3% yields a benefit-cost ratio of 3.84 and a discount rate of 5% yields a benefit-cost ratio of 2.27. The internal rate of return (ρ) can be obtained from solving the following equation: $\sum_{t=23}^{t=66} \frac{Y_t^{tr} - Y_t^{co}}{(1+\rho)^t} = C_{t=0}$, where Y_t^{tr} and Y_t^{co} are the benefits in treatment and control group and C are the program costs.

We use the organizational costs per child of 1,000 EUR. These costs do not comprise the opportunity costs incurred by mentors if they spend one afternoon per week on

activities other than studying or working. However, they also do not comprise long-term benefits incurred by mentors, e.g., in the form of later job prospects that may arise if employers react to the mentoring engagement as a positive signal of their prosociality (Heinz and Schumacher, 2017).

For the estimates presented in the text, we assume a persistent treatment effect of 20% of a standard deviation. Due to self-productivity and dynamic complementarity, the true long-term effect might be larger. Nonetheless, it might also be smaller if the program effect fades out over time (long-run prosociality differences by SES are reported in table B.3). Table B6 thus displays benefit-cost ratios and internal rates of return for different long-term program effects and discount rates. In all cases, program returns are positive and the internal rate of return varies between 5.5% and 9.6%.

Effect size	Benefit-cost ratio			IRR
	Discount rate			
	0.01	*0.03*	*0.05*	
0.25	8.64	4.81	2.83	0.096
0.2	6.91	3.84	2.27	0.085
0.15	5.18	2.88	1.70	0.055

Table B6: This table provides estimates of benefit-cost ratios and internal rates of return for different discount rates and program effect sizes.

We focus on labor market earnings, because they are easily quantifiable and of major interest to an economic audience. Benefits of more education, health and well-being are not estimated due to data limitations as are the effects on future generations. In this sense, our estimates are probably a lower bound of the true effect.

C Experimental protocols

(translations from German)

Binary dictator game

"Have a look at these paper stars I've got here."

⇒ Show the stars to the child.

"We now want to play a game in which you can win stars. Later on you can exchange these stars for a present. The more stars you win the better will the present be you are going to get afterwards."

⇒ Take the child to the boxes with the presents; explain that the presents become better/bigger going from box 1 to box 4 and that one needs more stars to get a better/bigger present.

"You get some stars and you can give stars to another child as well. Here I have two paper bags. One is for you. Let's write your name on it. The other bag is for another child that is of your age and also lives here in Bonn (Cologne). But both of us do not know exactly who that other child is. So we will not write a name on the second bag. The bags are for the stars that you and the other child are going to get in this game."

"Here are two sheets of paper."

⇒ Place both sheets of paper in front of the child (right and left).

"On each sheet you can see two circles with arrows. On both sheets one arrow is pointing to the bag of the other child and the other arrow is pointing to you and your bag. On the first sheet I place one star in the circle that is closer to the bag of the other child and I place one star in the circle that is closer to you."

⇒ Place the stars in the circles now.

"On the second sheet I place two stars in the circle that is closer to you. In the other circle that is pointing to the bag of the other child I place no star. You can now choose one of those sheets. If you choose this one, this arrow is pointing at you.

That means you are getting what is placed in this circle, one star."

⇒ Point to the respective circles.

"The other circle is pointing to the bag of the other child. This means that the other child is getting what is placed in this circle, one star. If you choose this sheet, you get two stars and the other child gets no star."

⇒ Point to the first sheet.

"If you choose this sheet, what will the other child get? And what will you get?"

⇒ Point to the second sheet.

"If you choose the second sheet, what will the other child get? And what will you get?"

Check of understanding of the rules

⇒ Repeat rules up to three times.

⇒ If the child has difficulties answering the control questions, explain the rules again. Ask the questions again. If the child does not understand the rules at all, play the game nonetheless so that the child will not be disappointed. Do not play the game if the child is frustrated due to the lack of understanding and does not want to play.

"Okay, which sheet do you choose?"

⇒ After the decision, put the stars into the respective bags. ⇒ **Remove both bags.** Place the bag of the participating child nearby. Put the bag of the other child away.

"I will make sure that the other child gets something nice for the stars."

Continuous dictator game A

"Now we will again play a game in which you can win stars. Later on you can exchange the stars for a present. The more stars you win, the better your present will be. You will get some stars and you can give stars to another child as well. Here I have two paper bags. The first one is for you. Let's write your name on it. The second bag is for another child that is of your age and lives in this area but not in Bonn (Cologne). Both of us do not know who that other child is. Therefore we do not write a name on the second bag. The bags are for the stars that you and the other child will get in this game."

⇒ Put both bags side by side on the table in front of the child.

"Look! Here are 6 stars."

⇒ Put the stars in front of the child between the two bags.

"Now you can decide how many stars you put on your bag and how many you put on the bag of the other child. The stars on your bag (point to the bag) are for you. The stars on the other bag are for the other child from another town nearby. I will make sure that the other child gets something nice for the stars.

You can decide how you want to divide the stars. You can split the stars or you can put all stars on one bag. How do you want to divide the stars? Now, please put all 6 stars on the bags in a way you would like to have it."

⇒ Child puts the stars on the bags.

"Okay. How many stars do you get? And how many stars does the other child get?"

⇒ In case the answers are not correct: explain the correct answer and ask for new suggestion. In case both answers are correct:

"Fine. Let's now put the stars in the bags."

⇒ Remove the bags.

Continuous dictator game B

"We will now play a similar game in which you can win stars. You will get some stars and you can give stars to another child again. Here are two bags. We will write your name on the bag that is for you. The other bag is for another child that is about your age. This child lives in Africa, which is very far away. The child cannot live with its parents, e.g., because they are too poor, ill or perhaps even dead. This child has only a few things to play with. We both do not know the name of the child therefore we do not write a name on the other bag. Again the bags are for the stars that you and the other child will receive."

⇒ Put both bags side by side on the table in front of the child.

"Look! Here are 6 stars."

⇒ Put the stars in front of the child between the two bags.

"Now you can decide how many stars you put on your bag and how many you put on the bag of the other child. The stars on your bag (point to the bag) are for you. The stars on the other bag are for the other child that lives in Africa without his or her parents and only with a few toys. I will make sure that the other child gets something nice for the stars.

You can decide how you want to divide the stars. You can split the stars or you can put all the stars on one bag. How do you want to divide the stars? Now, please put all 6 stars on the bags in a way you would like to have it."

⇒ Child puts the stars on the bags.

"Okay. How many stars do you get? And how many stars does the other child get?"

⇒ In case the answers are not correct: explain the correct answer and ask for new suggestion. In case both answers are correct:

"Fine. Let's now put the stars in the bags."

⇒ Remove the bags