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

Fabian Kosse

Ludwig-Maximilians-Universität München and Institute on Behavior and Inequality

Thomas Deckers

University of Bonn

Pia Pinger

University of Cologne and Institute on Behavior and Inequality

Hannah Schildberg-Hörisch

University of Düsseldorf

Armin Falk

Institute on Behavior and Inequality and University of Bonn

This study presents evidence on the role of social environment for the formation of prosociality. We show that socioeconomic status (SES) as well as intensity of mother-child interaction and mothers' prosocial attitudes are related to elementary school children's prosociality. We also present evidence on a randomly assigned variation of the social environment, providing children with a mentor for 1 year. Our data reveal a significant and persistent increase in prosociality in the treatment relative to the control group. Moreover, enriching the social environment

We thank the editor, James Heckman, and three anonymous referees. For comments and discussions, we are grateful to S. Alan, Z. Bašić, M. Bauer, A. Becker, T. Boneva, B. Born,

Electronically published January 2, 2020
[*Journal of Political Economy*, 2020, vol. 128, no. 2]

closes the prosociality gap between low- and high-SES children. A mediation analysis suggests that prosociality develops in response to prosocial role models and intense social interactions.

I. Introduction

Prosociality is a particularly important aspect of human personality and affects a wide range of economic decisions and outcomes—for example, the provision of public goods, contract enforcement, management of commons, governmental and judicial efficiency, and economic growth (Knack and Keefer 1997; La Porta et al. 1997; Fehr and Gächter 2002; Ostrom et al. 2002; Henrich et al. 2004; Guiso, Sapienza, and Zingales 2009; Cooper and Kagel 2015; 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 noncognitive skills (see, e.g., Heckman, Stixrud, and Urzua 2006). It is also reflected in economic theory, which, pioneered by Becker's (1976) work on altruism, incorporates other-regarding preferences and beliefs as key components (e.g., Rabin 1993; Fehr and Schmidt 1999, 2006; Charness and Rabin 2002; Bénabou and Tirole 2006; Falk and Fischbacher 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

A. Cappelen, F. Cunha, S. DellaVigna, T. Dohmen, J. Dovern, S. Durlauf, P. Eisenhauer, F. Englmaier, B. Enke, E. Fehr, C. Felfe, H.-M. von Gaudecker, H. Gerhardt, L. Götte, H. Hermes, J. Hermle, T. Kautz, M. Kocher, K. Kosse, S. Kube, P. Lergetporer, J. List, U. Malmendier, I. Rasul, S. Resnjanskij, C. Roth, M. Sandner, P. Schempp, D. Schunk, J. Schupp, S. Schurer, U. Sunde, M. Sutter, U. Volz, J. Winter, L. Wößmann, F. Zimmermann, U. Zölitz, and many seminar and conference participants. We thank M. Antony for great support in setting up the briq family panel and A. Bohlender, A. Jänsch, I. Leven, K. Thümmel, and all interviewers from Taylor Nelson Sofres/Kantar, as well as D. Esch, H. Krauß, and H. Müller-Kohlenberg from Balu und Du for their extraordinary effort. Financial support through the Leibniz Programme of the German Science Foundation, the Benckiser Stiftung Zukunft, and the German Research Foundation through Collaborative Research Center Transregio 224 is gratefully acknowledged. Data are provided as supplementary material online.

[1] Using data from the German Socio-Economic Panel (SOEP), we find, e.g., that a 1 standard deviation increase in prosociality translates into a 0.45 standard deviation improved subjective well-being and a 2 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; tables A1–A9, B1–B6 are available online).

social environment. It documents that prosociality in children is malleable and shows 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). To obtain exogenous variation in these stimuli, we randomly assigned a sample of elementary school children to an enriched social environment in the form of a mentoring program. This mentoring program is a well-established nonprofit program called Balu und Du (German for "Baloo and You"). It provides children with a volunteer mentor for the duration of 1 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 or 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 socioeconomic background information, interaction patterns, and assessments of personality regarding their children and themselves. 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 Strengths and Difficulties Questionnaire (SDQ; Goodman 1997). Our measures thus provide a broad-ranging characterization of prosocial disposition

[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).

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 socioeconomic status (SES) and treatment status. Using the sociodemographic 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 a 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 1-year treatment period was closely followed by a posttreatment wave of interviews (wave 2) and a 2-year follow-up wave (wave 3). In the main analysis we focus on posttreatment wave 2 data, but we also use wave 1 data 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 the 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 scored 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 per child per year).[3] Moreover, the observed high- to low-SES developmental gap in prosociality is closed; that is, 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 and, most importantly, persistence of the high- to low-SES gap and the treatment effect.[4] 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 for 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

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

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

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

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 (e.g., Hawkins et al. 1999, 2005) 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 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 this paper is organized as follows. In the next section, we discuss recruitment, our measures, details of the mentoring intervention, and that intervention's implementation. Section III first presents descriptive evidence and deduces more specific research questions regarding the relationship between a child's social environment and the formation of prosociality. Subsequently, it unveils the causal effect of enriching a child's social environment on prosociality and delineates the underlying mechanisms. Section IV concludes.

II. Study Design

This section introduces the design of the briq family panel (BFP), which is the base of this 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. 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.

A. Recruitment of Sample

Figure 1 presents a flowchart of the timing, sampling, and procedural details of the BFP. Recruiting started in summer 2011. We used official registry data to obtain more than 95% of the addresses of families (with children aged from 7 to 9) living in the German cities of Bonn and Cologne.

[6] A 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 John and Thomsen (2015), Cappelen et al. (2016), Alan and Ertac (2017), and Rao (2019).

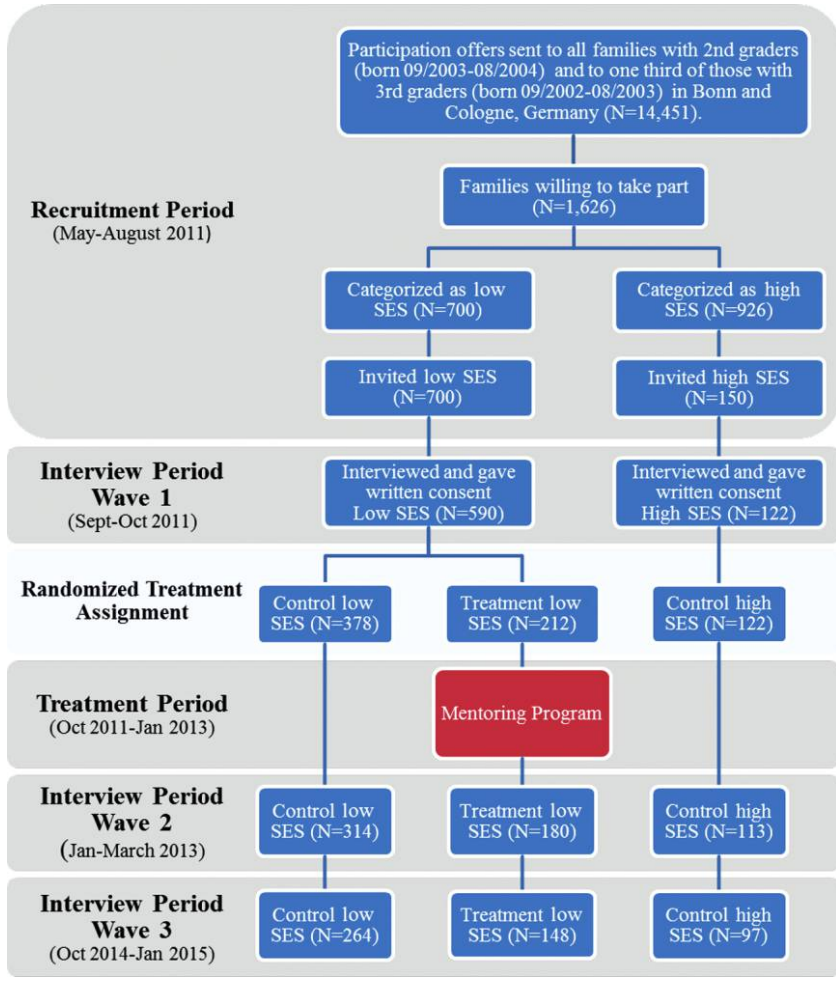

FIG. 1.—Flowchart of sampling and procedural details.

Invitations to take part in the mentoring program and the interviews were sent to all families with children born between September 2003 and August 2004 and to one-third of the families with children born between September 2002 and August 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 of taking part in the mentoring program and the interviews. We

[7] Nearly all the 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.

informed parents that participation in the mentoring program was not guaranteed because of limited capacity (which was the case). We asked them to send back a short questionnaire concerning socioeconomic characteristics of the household and to sign a nonbinding 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. SES reflects the level of resources available at the household level—that is, material, educational, and time resources. Accordingly, a household was classified as low SES if at least one of the three following criteria was met: (1) low income (equivalence income of the household is lower than €1,065; this corresponds to the 30% quantile of the German income distribution);[9] (2) low education (neither the mother nor the father of the child has a school-leaving degree qualifying for university studies); or (3) single-parent status (a parent is classified as a single parent if he or 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. To be eligible for treatment, they had to participate in the first wave of interviews (fall 2011) and provide written consent to allow for 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 14 subgroups 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

[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 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 greater share of children in Bonn to the ITT group. Therefore, assignment into treatment was random, conditional on city of residence. However, conditioning on city of residence does not affect our results (cf. fig. 2 and table A1).

transmitted contact information of the ITT group to the mentoring organization that initiated the treatment. 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 1-year treatment period, all families who had participated in wave 1 were invited to take part in the posttreatment wave (wave 2); 85.3% (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 2-year follow-up.[12]

B. 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 app. C; apps. A–C are available online), which was identical across all three waves. Overall, interviews lasted about 1 hour. For participation in the interview, mothers received €35 in wave 1 and €45 in waves 2 and 3.

We took care to create a pleasant, nonstressful interview situation by seating a mother and her child in the same room. However, 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, socioeconomic background of the family, and details on how parents spend time with the child, including joint activities, as well as economic and social preferences of the mother.

[12] See sec. III.D for a discussion of baseline balance and attrition.

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 fig. A2; figs. A1–A7 are available online). 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 reached an age where toys are no longer appropriate as an incentive. We therefore changed the incentive to money in this wave, with one star corresponding to €0.30.

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: (1) mentors received no information about the elicited measures, to avoid any form of "training to the test"; (2) 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); (3) the intervention was not mentioned during the data-collection phase; and (4) the research team never interacted directly with the children or their parents.

C. Measures of Prosociality and Social Interaction

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 various 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, and our measure of social interaction patterns.

1. 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- or herself and another anonymously matched child of the same age. In particular, we conducted one binary

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

dictator game and two continuous versions of dictator games with varying receivers (for experimental protocols, see app. C).

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 stars (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. 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 or her parents because they either are "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 the form of toys. 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 "an African country," we collaborated with an SOS Children's Village in Togo.

2. 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 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. 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).

3. 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 SDQ (Goodman 1997). The SDQ is a well-established behavioral screening survey comprising five subscales. The prosocial scale includes five items regarding the child, which read as follows:[14] "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).

4. 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. 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,

[14] We used the wording of the official German SDQ version; see https://www.sdqinfo.org.

the joint measure is calculated as the equally weighted score of the two observed facets.

5. Prosociality of Mothers and Mentors

To obtain measures of prosociality for mothers and mentors, respectively, we proceed as similarly to our approach for the children as possible: 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 app. sec. B.2). Trust was measured using the two items "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 11-point Likert scale. Other-regarding behavior was measured using the "Big Five" dimension "agreeableness" in the form of a three-item version (seven-point Likert scale; Lang et al. 2011). The items, given as measures of how one sees oneself, read as follows: "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 wording similar to 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 standardized measures as our measure of prosociality for mothers and mentors.

6. Social Interaction Patterns

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 per afternoon, with a mean (of all groups) 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 94.8% and thus much higher (see also app. sec. B.5).

D. Mentoring Intervention

The intervention that we implemented is a well-established nonprofit mentoring program called Balu und Du. In this program, elementary school children are provided with a mentor for the duration of 1 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 or 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 or she is valued on the basis that the mentor regularly spends time with him or 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 biweekly monitoring meetings where mentors receive suggestions for activities with the mentored child and discuss potential problems. To date, the Balu und Du 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 (fig. A1). Variation in duration is mainly due to unforeseeable events, such as moving decisions of parents or mentors because of a job change.

[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 by the federal government of North Rhine–Westphalia (Germany) in 2006. More details about the mentoring program can be found at https://www.balu-und-du.de and in an overview article by Müller-Kohlenberg and Drexler (2013).

On average, treated children met their mentor 22.8 times (11.9 SD), 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 because of a local shortage of mentors, mentor refusals, or coordination problems between 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 we discuss local average treatment effects (LATEs) in section III.D.

III. 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 children. Section III.B then unveils the causal effect of enriching a child's social environment on prosociality, and section III.C presents a mediation analysis concerning the underlying mechanisms and evinces how prosociality develops in elementary school children. Finally, section III.D provides numerous robustness checks. This includes tests for baseline balance and 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, SD = 0.48). However, to study long-run effects of the intervention and persistence of parental background over a 2-year developmental period, we also refer to wave 3 data.

A. 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 (Benenson, Pascoe, and Radmore 2007; Bauer, Chytilová, and Pertold-Gebicka 2014; Falk and Kosse 2016). 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

TABLE 1
PROSOCIALITY AND PARENTAL BACKGROUND

	PROSOCIALITY OF CHILDREN (Standardized)			
	(1)	(2)	(3)	(4)
High SES	.226**			.140
	(.106)			(.105)
Prosociality of mother (standardized)		.258***		.228***
		(.048)		(.048)
Social interaction (standardized)			.214***	.167**
			(.048)	(.049)
Constant	−.127**	−.083*	−.064	−.117**
	(.059)	(.048)	(.048)	(.056)
Observations	418	418	418	418
R^2	.010	.067	.042	.098

NOTE.—Prosociality measures of mother and child are constructed as equally weighted scores of standardized measures of altruism, trust, and other-regarding behavior. Prosociality and social-interaction measures are standardized, and high SES is a dummy. For reasons of comparability, the sample is restricted to individuals with nonmissing values in all used variables. Data come from wave 2. The sample consists of control low SES and control high SES. Coefficients are ordinary least squares estimates, with White robust standard errors shown in parentheses.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

dummy variable for high SES (see col. 1) reveals that children from high-SES households are substantially and significantly more prosocial than those from low-SES households ($p < .05$). 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. 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 socioeconomic disparities in child prosociality? Informed by the literature in developmental psychology, we identify and study two important drivers: prosocial attachment figures and intense

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

[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, and 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 ($p = .737$, $N = 418$, two-sided t-test) nor specifically Turkish background ($p = .695$, $N = 418$, two-sided t-test) is related to children's prosociality.

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. High-SES children live with adults who are themselves more prosocial. On average, high-SES mothers score 15.6% of a standard deviation higher on prosociality than low-SES mothers in our data ($p = .160$, $N = 418$, 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 ($p < .01$, $N = 418$, 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 1 standard deviation increase in a mother's prosociality is related to an increase of 25.8% of a standard deviation in her child's prosociality ($p < .01$).[19] Moreover, column 3 shows that spending more time on intense social interactions positively relates to children's prosociality, as children who experience a 1 standard deviation higher intensity of social interaction with their mothers are 21.4% of a standard deviation more prosocial ($p < .01$).

In column 4 of 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 about 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: (1) Is prosociality malleable in elementary school children, and can it be enhanced by enriching a child's social environment? (2) 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.

[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); Zumbuehl, Dohmen, and Pfann (2013); and Alan et al. (2017).

B. *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 this relationship is causal. Children participating in the mentoring program are randomly assigned to an enriched social environment in the 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.

Figure 2*A* shows our first main result. In wave 2 (posttreatment), 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 ($p < .01$, $N = 494$, two-sided t-test). Figure 2 also indicates that the high- to low-SES developmental gap in prosociality is closed for children in the

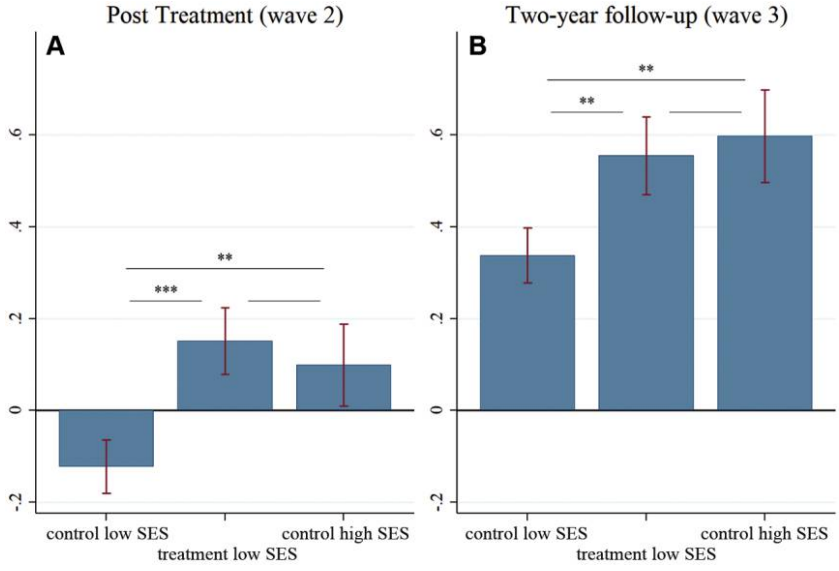

FIG. 2.—Significantly higher levels of prosociality for treated children compared with untreated children (treatment low SES vs. control low SES). There is no significant difference between treatment low SES and control high SES. The scale on the *Y*-axis indicates *z*-scores (i.e., standardized measures) of children's prosociality. Standardization is conducted using the distribution of wave 2. Error bars show standard errors of the means. Three asterisks and two asterisks indicate significant differences at the 1% and 5% level, respectively (two-sided *t*-tests). *A*, Treatment low SES versus control low SES: $N = 494$; control high SES versus control low SES: $N = 427$; treatment low SES versus control high SES: $N = 293$. *B*, Treatment low SES versus control low SES: $N = 411$; control high SES versus control low SES: $N = 358$; treatment low SES versus control high SES: $N = 243$.

treatment group: children from treatment low SES and control high SES score very similarly on prosociality ($p = .651$, $N = 293$, 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 ($p < .05$), trust ($p = .05$), and other-regarding behavior ($p = .266$; 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, for example, experimental measures that are based on revealed preferences, especially in disadvantaged samples (Borghans et al. 2008; Sandner and Jungmann 2016). In fact, if this subscale is omitted from the overall prosociality score, our results increase in size and significance (see col. 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 ($p < .05$), indicating that the treatment effect increases by 21.4% of a standard deviation of prosociality if the mother scores 1 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 ($p < .05$) if the child experiences 1 standard deviation less intensive social interaction in his or her home environment.

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, for example, 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 underscore the potential of childhood interventions in general. To investigate this issue, we use the 2-year follow-up data (wave 3); see figure 2*B*.

The figure illustrates three important findings. First, over the time span of 2 years, we observe a general increase in prosociality in our sample of elementary school children. The average increase amounts to more than 40% of a wave 2 standard deviation and holds for all groups under

[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; $p = .010$, $F = 3.78$, $N = 491$).

study—that is, irrespective of SES and treatment.[21] To the best of our knowledge, this is the first documentation of an increase in prosociality within 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; Bašić, Falk, and Kosse 2019). Second, the high- to 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 25.9% of a wave 2 standard deviation ($p < .05$, $N = 358$, two-sided t-test).[22] Third, and most importantly, the treatment effect reported in figure 2*A* proves to be remarkably robust over time. As shown in figure 2*B*, 2 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; $p < .05$, $N = 411$, two-sided t-test). Consequently, prosociality in wave 3 does not significantly differ between treatment low-SES and control high-SES children ($p = .749$, $N = 243$, two-sided t-test). In other words, the developmental gap that was closed in response to treatment remains closed more than 2 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.

C. Mechanisms

In this subsection, 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.

Following the approach and notation laid out in Heckman, Pinto, and Savelyev (2013), we focus on a linear production function and write child

[21] A test for differences in posttreatment 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 ($p = .611$).

[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).

i's prosociality, when the program assignment is set to "treated" ($d = 1$) or "control" ($d = 0$), as

$$P^C_{i,d} = \kappa_d + \alpha^P_d P^A_{i,d} + \alpha^I_d I^A_{i,d} + \alpha^U_d U_{i,d} + \beta_d X_i + \epsilon_{i,d}, \quad d \in \{0, 1\}, \quad (1)$$

where $P^C_{i,d}$ represents counterfactual prosociality of child i, κ_d is an intercept, and α^P_d and α^I_d are scalar parameters, denoting the effects of prosocial attachment figures and social interactions on child prosociality. Moreover, α^U_d is a vector denoting the effect of several unobserved factors ($U_{i,d}$) on child prosociality, and β_d is a vector that captures the effect of preprogram 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 ($P^A = P^M$),[23] and I^A is measured as the average share of highly interactive activities experienced per afternoon. Children in the treatment group who interacted with a mentor are influenced by both 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 the 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 or her mother (M) or mentor (B), respectively (for details, see app. sec. B.5).[25] The value of 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 activities per afternoon, taking into account the fact that treated children spend up to one afternoon per week with their mentors.

[23] Prosociality was measured only in mothers and not in fathers. However, Dohmen et al. (2012), e.g., 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 the relationship, see MacCallum and Beltman (2002).

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

There exists substantial heterogeneity in the share of highly interactive activities conducted with both mothers and mentors, as documented in the respective time-use questionnaires (see app. sec. 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 ($p < .01$, two-sided t-test). The reason is that the self-selected volunteer mentors are on average more than 1 standard deviation more prosocial than the low-SES mothers in our sample (see table B4). Figure A5 also shows that the ITT children engage in on average 14% more intense social activities per afternoon (40% in terms of standard deviations; $p < .01$, 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

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

where $\tau = \kappa_d + \Sigma_{j \in J_U} \alpha_d^j \mathbb{E}[U_d^j]$, 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, either individually or 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 semiparametric identification of the ITT effect has been discussed in Yamamoto (2013) and Keele, Tingley, and Yamamoto (2015). For identification that relies on instrumental variables for treatment compliance and mediators, see Frölich and Huber (2017).

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] The top bar 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 middle and bottom bars 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 remain similar, indicating that both inputs affect child prosociality separately and independently.

D. 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-differences 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 on the basis of 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 data set 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 (for details, see app. sec. B.4.1).[30] The reweighted ITT effects and SES

[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 sec. III.A, 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.

[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).

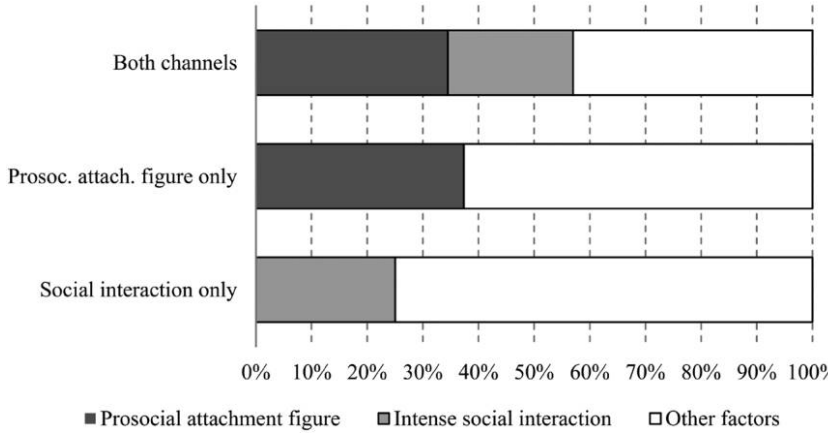

FIG. 3.—Estimates of the decomposition as described in equation (2). Each bar represents the total treatment effect (wave 2) normalized to 100%. The dark gray area displays the share of the treatment effect explained by changes in prosociality of attachment figures. The light gray area displays the share of the treatment effect explained by the increase in intense social interactions. The white area indicates unmeasured other factors. The top bar 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: $p < .01$ for all channels in all three decompositions. For details on the decomposition, see Heckman, Pinto, and Savelyev (2013).

gaps are very similar to the estimated main effects (see fig. A6), except for a slight increase in the ITT effect (see also 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 (i.e., the probability that households did not participate in waves 2 and 3 after having participated in wave 1) are not related to pretreatment prosociality, treatment assignment, or their interaction (see table A4). Moreover, reweighting the observed data using inverse probabilities of participation in the wave 2 and 3 interviews confirms our findings reported in section III.B (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 on the basis of extreme assumptions about selection. The results shown in table A6 are in line with the previously reported results.

[31] The predicted probabilities result from a Probit model of a binary selection indicator (indicating whether wave 2 or 3 interviews were conducted) as a function of pretreatment prosociality and treatment assignment.

The pretreatment measure of prosociality can further be used to investigate baseline imbalance. The randomization procedure was successful, as pretreatment prosociality does not differ by treatment status ($p = .662$, $N = 590$, two-sided t-test). Moreover, we can use the pretreatment measure of prosociality to obtain difference-in-differences estimates as a robustness check of our main estimates. In sum, and consistent with baseline balance and nonselectivity of attrition, the difference-in-differences 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 because of a shortage of mentors or could not be realized because of moving, other coordination problems, or refusals by mentors or families.[32] We thus focus on ITT estimates since they rely on only 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 LATEs.[33] In table A8, we present two-stage least squares estimates using the random assignment as an instrument for actual treatment. Because control group children could not participate in the program (by design), the LATE equals the average treatment effect on the treated, amounting to 39.8% of a standard deviation. Thus, in line with a matching 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

[32] The complier children do not significantly differ from the noncompliers in any relevant dimension, such as initial child prosociality, mother prosociality, parental investments, etc. Participant selection on program gains for the majority of noncompliers 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?"

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 app. sec. B.7). Nevertheless, it is comforting that the treatment effect on prosociality is still significantly different from zero at the 1% level (for adjusted p-values, see table B5) even after we conduct a multiple-hypothesis 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 genders of child and mentor and the addresses of families and mentors were known to Balu und Du. Male mentors were matched only 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 app. sec. B.5) and the increase in children's prosociality (between waves 1 and 2). The results are displayed in table A9. Regarding the gender composition, we find that same-sex matches tend to result in a better relationship quality ($p < .1$, two-sided t-test)[36] and a slightly larger increase in children's prosociality ($p = .507$, two-sided t-test). We also find that matches are more beneficial if mentors and mentees live nearby in terms of commuting distance ($p = .632$ for relationship quality, and $p < .05$ for the increase in prosociality; two-sided t-test). We do not observe similar effects for commonalities in parental background.

[34] Further, note that intensity of social interaction as described in sec. II.C is also not significantly related to risk-taking ($p = .346$) or patience ($p = .753$).

[35] For a portrayal of the algorithm, see Heckman et al. (2010a) and Romano and Wolf (2016).

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

IV. 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 is remarkably robust over time: the positive treatment effect of the enriched social environment on prosociality is enduring and observed 2 years after the end of the intervention. It is based on a comprehensive measure of prosociality—consisting of 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, 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

[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 level of prosociality (0.2), the average number of yearly hours worked in Germany (about 1,400; 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 app. sec. 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.

we study in this paper may be socially desirable because they close developmental gaps arising from socioeconomic 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 greatest 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 or her apart from the child's usual caregivers. In terms of policy, 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), for example, regarding the reduction of antisocial behaviors, conduct problems, or peer relationship difficulties. Last, our results point toward 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 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 rely exclusively 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): 856–61.

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

Alan, Sule, and Seda Ertac. 2017. "Belief in Hard Work and Altruism: Evidence from a Randomized Experiment." HCEO Working Paper no. 2017-053, Human Capital and Econ. Opportunity Global Working Group, Chicago.

———. 2018. "Fostering Patience in the Classroom: Results from a Randomized Educational Intervention." *J.P.E.* 126 (5): 1865–911.

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." Working paper, Sci. Po, Paris.

Almås, Ingvild, Alexander W. Cappelen, Erik Ø. Sørensen, and Bertil Tungodden. 2010. "Fairness and the Development of Inequality Acceptance." *Science* 328 (5982): 1176–78.

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.

Bašić, Zvonimir, Armin Falk, and Fabian Kosse. 2019. "The Development of Egalitarian Norm Enforcement in Childhood and Adolescence." *J. Econ. Behavior and Org.*, forthcoming.

Bauer, Michal, Julie Chytilová, and Barbara Pertold-Gebicka. 2014. "Parental Background and Other-Regarding Preferences in Children." *Experimental Econ.* 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." *Ann. Rev. Econ.* 4 (1): 453–78.

Becker, Gary S. 1976. "Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology." *J. Econ. Literature* 14 (3): 817–26.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *A.E.R.* 96 (5): 1652–78.

Benenson, Joyce F., Joanna Pascoe, and Nicola Radmore. 2007. "Children's Altruistic Behavior in the Dictator Game." *Evolution and Human Behavior* 28 (3): 168–75.

Bisin, Alberto, and Thierry Verdier. 2000. "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits." *Q.J.E.* 115 (3): 955–88.

———. 2001. "The Economics of Cultural Transmission and the Dynamics of Preferences." *J. Econ. 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." *J. Human Resources* 43 (4): 972–1059.

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

Boyd, Robert, Peter J. Richerson, and Joseph Henrich. 2011. "The Cultural Niche: Why Social Learning Is Essential for Human Adaptation." *Proc. Nat. Acad. Sci. USA* 108 (suppl. 2): 10918–25.

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 Working Paper no. 9767, Inst. Study Labor, Bonn.

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): 1478–85.

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

Carpenter, Jeffrey, and Erika Seki. 2011. "Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyama Bay." *Econ. Inquiry* 49 (2): 612–30.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Q.J.E.* 125 (3): 1101–44.

Case, Anne, Darren Lubotsky, and Christina Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient." *A.E.R.* 92 (5): 1308–34.

Cesarini, David, Christopher T. Dawes, James H. Fowler, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace. 2008. "Heritability of Cooperative Behavior in the Trust Game." *Proc. Nat. Acad. Sci. USA* 105 (10): 3721–26.

Cesarini, David, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace. 2009. "Genetic Variation in Preferences for Giving and Risk Taking." *Q.J.E.* 124 (2): 809–42.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Q.J.E.* 117 (3): 817–69.

Cooper, David, and John H. Kagel. 2015. "Other-Regarding Preferences: A Selective Survey of Experimental Results." In *Handbook of Experimental Economics*, vol. 2, edited by John H. Kagel and Alvin E. Roth, 217–89. Princeton, NJ: Princeton Univ. Press.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *A.E.R.* 97 (2): 31–47.

Currie, Janet, and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *Q.J.E.* 118 (4): 1495–532.

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 Working Paper no. 11158, Inst. Study Labor, Bonn.

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market." *Q.J.E.* 132 (4): 1593–640.

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

———. 2012. "The Intergenerational Transmission of Risk and Trust Attitudes." *Rev. Econ. Studies* 79 (2): 645–77.

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 Econ.* 45 (1): 40–58.

Duncan, Greg J., and Katherine Magnuson. 2013. "Investing in Preschool Programs." *J. Econ. Perspectives* 27 (2): 109–32.

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." *J. Vocational Behavior* 72 (2): 254–67.

Eisenberg, Nancy, Richard A. Fabes, and Tracy L. Spinrad. 2006. "Prosocial Development." In *Social, Emotional, and Personality Development*, vol. 3 of *Handbook of Child Psychology*, edited by Nancy Eisenberg, 646–718. Hoboken, NJ: Wiley.

Eisenberg, Nancy, Tracy L. Spinrad, and Ariel Knafo-Noam. 2015. "Prosocial Development." In *Socioemotional Processes*, vol. 3 of *Handbook of Child Psychology and Developmental Science*, edited by Michael E. Lamb, 610–56. Hoboken, NJ: Wiley.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *Q. J.E.* 133 (4): 1645–92.

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 Working Paper no. 9674, Inst. Study Labor, Bonn.

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

Falk, Armin, and Fabian Kosse. 2016. "Early Childhood Environment, Breastfeeding and the Formation of Preferences." HCEO Working Paper no. 2016-036, Human Capital and Econ. Opportunity Global Working Group, Chicago.

Fehr, Ernst, Helen Bernhard, and Bettina Rockenbach. 2008. "Egalitarianism in Young Children." *Nature* 454 (7208): 1079–83.

Fehr, Ernst, and Urs Fischbacher. 2003. "The Nature of Human Altruism." *Nature* 425 (6960): 785–91.

Fehr, Ernst, Urs Fischbacher, Bernhard von Rosenbladt, Jürgen Schupp, and Gert G. Wagner. 2002. "A Nation-Wide Laboratory." *J. Appl. Soc. Sci. Studies* 122:519–42.

Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415 (6868): 137–40.

Fehr, Ernst, Daniela Glätzle-Rützler, and Matthias Sutter. 2013. "The Development of Egalitarianism, Altruism, Spite and Parochialism in Childhood and Adolescence." *European Econ. Rev.* 64 (1): 369–83.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Q. J.E.* 114 (3): 817–68.

———. 2006. "The Economics of Fairness, Reciprocity and Altruism—Experimental Evidence and New Theories." In *Handbook of the Economics of Giving, Altruism and Reciprocity*, vol. 1, edited by Serge-Christophe Kolm and Jean Mercier Ythier, 615–91. Amsterdam: North-Holland.

Foster, Lisa. 2001. "Effectiveness of Mentor Programs: Review of the Literature from 1995 to 2000." Tech. report, California Res. Bur., California State Library, Sacramento.

Frölich, Markus, and Martin Huber. 2017. "Direct and Indirect Treatment Effects—Causal Chains and Mediation Analysis with Instrumental Variables." *J. Royal Statis. Soc. B* 79 (5): 1645–66.

Goodman, Robert. 1997. "The Strengths and Difficulties Questionnaire: A Research Note." *J. Child Psychology and Psychiatry* 38 (5): 581–86.

Grossman, Jean Baldwin, and Joseph P. Tierney. 1998. "Does Mentoring Work? An Impact Study of the Big Brothers Big Sisters Program." *Evaluation Rev.* 22 (3): 403–26.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Cultural Biases in Economic Exchange?" *Q. J.E.* 124 (3): 1095–131.

Harbaugh, William T., and Kate Krause. 2000. "Children's Altruism in Public Good and Dictator Experiments." *Econ. 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 Pediatrics and Adolescent Medicine* 153 (3): 226–34.

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 Pediatrics and Adolescent Medicine* 159 (1): 25–31.

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

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

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 Savelyev, and Adam Yavitz. 2010a. "Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program." *Quantitative Econ.* 1 (1): 1–46.

———. 2010b. "The Rate of Return to the HighScope Perry Preschool Program." *J. Public Econ.* 94 (1/2): 114–28.

Heckman, James J., and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Ann. Rev. Econ.* 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." *A.E.R.* 103 (6): 2052–86.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *J. Labor Econ.* 24 (3): 411–82.

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: Oxford Univ. Press.

Hertwig, Ralph, and Andreas Ortmann. 2001. "Experimental Practices in Economics: A Methodological Challenge for Psychologists?" *Behavioral and Brain Sci.* 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 Econ. Behavior* 94:122–41.

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, Org. Econ. Cooperation and Development, Paris.

Keele, Luke, Dustin Tingley, and Teppei Yamamoto. 2015. "Identifying Mechanisms behind Policy Interventions via Causal Mediation Analysis." *J. Policy Analysis and Management* 34 (4): 937–63.

Knack, Stephen, and Philip Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Q. J.E.* 112 (4): 1251–88.

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." *J. Human Resources* 54 (2): 430–67.

Kosfeld, Michael, Markus Heinrichs, Paul J. Zak, Urs Fischbacher, and Ernst Fehr. 2005. "Oxytocin Increases Trust in Humans." *Nature* 435 (7042): 673–76.

Kosse, Fabian, and Friedhelm Pfeiffer. 2012. "Impatience among Preschool Children and their Mothers." *Econ. Letters* 115 (3): 493–95.

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 Res. Methods* 43 (2): 548–67.

La Porta, Rafael, Florencio Lopez-de-Silane, Andrei Shleifer, and Robert W. Vishny. 1997. "Trust in Large Organizations." *A.E.R.* 87 (2): 333–38.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Rev. Econ. Studies* 76 (3): 1071–102.

MacCallum, Judith, and Susan Beltman. 2002. "Role Models for Young People: What Makes an Effective Role Model Program." Report, Nat. Youth Affairs Res. Scheme, Australian Clearinghouse Youth Studies, Hobart.

McCord, Joan. 1978. "A Thirty-Year Follow-Up of Treatment Effects." *American Psychologist* 33 (3): 284–89.

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, Hildegard, and Sibylle Drexler. 2013. "Balu und Du ('Baloo and You')—A Mentoring Program: Conception and Evaluation Results." In *Mentoring: Practices, Potential Challenges and Benefits*, edited by Michael F. Shaughnessy, 107–23. New York: Nova.

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): 1937–40.

OECD (Organization for Economic Cooperation and Development). 2017. "OECD.Stat." https://stats.oecd.org.

Ostrom, Elinor, Thomas Dietz, Nives Dolsak, Paul C. Stern, Susan Stonich, and Elke U. Weber. 2002. *The Drama of the Commons*. Washington, DC: Nat. Acad.

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

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *A.E.R.* 83 (5): 1281–302.

Rao, Gautam. 2019. "Familiarity Does Not Breed Contempt: Generosity, Discrimination and Diversity in Delhi Schools." *A.E.R.* 109 (3): 774–809.

Rodríguez-Planas, Núria. 2012. "Mentoring, Educational Services, and Incentives to Learn: What Do We Know about Them?" *Evaluation and Program Planning* 35 (4): 481–90.

Romano, Joseph P., and Michael Wolf. 2005. "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing." *J. American Statis. Assoc.* 100 (469): 94–108.

———. 2016. "Efficient Computation of Adjusted *p*-Values for Resampling-Based Stepdown Multiple Testing." *Statis. and 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?" *Econ. Letters* 141 (1): 73–76.

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

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." *J. Royal Statis. Soc. A* 174 (2): 369–86.

Sutter, Matthias, and Martin G. Kocher. 2007. "Trust and Trustworthiness across Different Age Groups." *Games and Econ. Behavior* 59 (2): 364–82.
Tierney, Joseph P., and Jean Baldwin Grossman. 2000. *Making a Difference: An Impact Study of Big Brothers Big Sisters.* Philadelphia: Public/Private Ventures.
United States General Accounting Office. 1991. "Discount Rate Policy." Washington, DC: Government Printing Office.
Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp. 2007. "The German Socio-economic Panel Study (SOEP)—Scope, Evolution and Enhancements." *J. Appl. Soc. Sci. Studies* 127 (1): 139–69.
Werner, Emmy E. 2000. "Protective Factors and Individual Resilience." In *Handbook of Early Childhood Intervention*, edited by Jack P. Shonkoff and Samuel J. Meisels, 115–32. Cambridge: Cambridge Univ. Press.
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." *J. Experimental Child Psychology* 114 (4): 543–50.
Yamamoto, Teppei. 2013. "Identification and Estimation of Causal Mediation Effects with Treatment Noncompliance." Working paper, Dept. Polit. Sci., MIT.
Zumbuehl, Maria, Thomas Dohmen, and Gerard Pfann. 2013. "Parental Investment and the Intergenerational Transmission of Economic Preferences and Attitudes." IZA Working Paper no. 7476, Inst. Study Labor, Bonn.