Individual Heterogeneities, Social Environment and Life Outcomes

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von

Fabian Kosse
aus Freiburg im Breisgau

Bonn 2015

Dekan:	Prof. Dr. Rainer Hüttemann
Erstreferent:	Prof. Dr. Armin Falk
Zweitreferent:	Prof. Dr. Thomas Dohmen

Tag der mündlichen Prüfung: 27.02.2015

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn (http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert.

Für meine Eltern

Acknowledgments

Many people supported me during the process of preparing this thesis. First and foremost, I want to thank Armin Falk and Thomas Dohmen who have been great supervisors. They provided very detailed and constructive advice at all stages of my dissertation project. Working together with them was a great inspiration. My sincere thanks are also due to Pia Pinger for valuable comments and for acting as head of my thesis committee.

Moreover, I thank my coauthors Anke Becker, Thomas Deckers, John Eric Humphries and Hannah Schildberg-Hörisch for their effort and patience. Special thanks are owed to Friedhelm Pfeiffer who supported my first steps in economic research and introduced me to the topic of non-cognitive skills. I have also greatly benefited from discussions with and comments by Sascha Alavi, Steffen Altmann, Markus Antony, Michal Bauer, Zvonimir Bašić, Michael Böhm, Jörg Breitung, Philipp Eisenhauer, Benjamin Enke, Hans-Martin von Gaudecker, Holger Gerhardt, Andreas Grunewald, James Heckman, Simon Jäger, Tobias Klein, Andreas Kleiner, Martin Kocher, Sebastian Kube, Philipp Lergetporer, Jonas Radbruch, Martin Reuter, Daniel Schunk, Frederik Schwerter, Marcel Stierl, Sabrina Strang, Nora Szech, Bernd Weber, Matthias Wibral, Tom Zimmermann and Maria Zumbühl. I further thank Hildegard Müller-Kohlenberg, Dominik Esch and Heiko Krauß for making Chapter 4 possible.

Special thanks also go to Urs Schweizer for developing the Bonn Graduate School of Economics (BGSE) into a stimulating and thriving research environment and to Silke Kinzig, Pamela Mertens and Britta Altenburg for making the BGSE run smoothly. Financial support by the German Research Foundation (DFG) through the BGSE is gratefully acknowledged. I further thank Birgit Jendrock and Stefanie Sauter of the Institute for Applied Microeconomics (IAME) for providing various kinds of support and always having an answer.

Being a part of the BGSE cohort 2009 was an amazing, extraordinary and unforgettable experience. I thank Anke Becker, Markus Behn, Carsten Dahremöller, Thomas Deckers, Inga Deimen, Tilman Drerup, Gerrit Frackenpohl, Christian Hilpert, Felix Ketelaar, Florian Kirsch, Harun Mirza, Christoph Roling, Ronald Rühmkorf, Paul Schempp, David Schönherr and Lidia Storjohann for being my colleagues and friends.

Ich danke meinen Eltern Claudia und Franz für alles, was sie in den letzten 30 Jahren für mich getan haben, und dafür, dass sie immer für mich da sind. Ich danke meiner Schwester Christin dafür, meine Projekte mit den Augen einer Naturwissenschaftlerin zu sehen, kritisch zu hinterfragen und mit mir zu diskutieren.

I am greatly indebted to Katherina Bode for her love and unconditional support. This thesis would not exist without her!

Contents

List of Figures

List of Tables

Introduction

Individuals differ not only by gender and hair color, they also exhibit heterogeneities in various probably not fully genetically determined dimensions as personality, preferences and skills.[1] Moreover, individuals live in different social environments and yield unequal life outcomes as health or income. Understanding how differences in these aspects affect each other is of relevance not only for behavioral sciences, but also for informing policy as it helps to uncover reasons for social mobility and to target intervention programs. In this regard, especially the understanding of malleability of individual heterogeneities and life outcomes in response to social environment is of great interest. In general, despite their fundamental importance, little is known about the interactions between the three aspects named in the title.

This dissertation consists of five self-contained chapters which jointly seek to contribute to a better understanding of the interactions within the triangle of individual heterogeneities, social environment and life outcomes. To do so, Chapter 1 and 2 consider the relation between individual heterogeneities, as personality and preferences, and life outcomes, as income, health and education. Chapter 3 and 4 analyze the role of (early) social environment on the development process of personality and preferences. Finally, Chapter 5 considers the direct effect of social environment on life outcomes.

Although both economists and psychologists seek to identify determinants of heterogeneities in behavior and life outcomes, they use different concepts to capture them. In Chapter 1, we first analyze the extent to which economic preferences and psychological concepts of personality, such as the Big Five and Locus of Control, are

[1] Providing evidence for this claim is part of Chapter 4

related.[2] We analyze data from incentivized laboratory experiments and representative samples and find only low degrees of association between economic preferences and personality. We then regress life outcomes (such as labor market success, health status, and life satisfaction) simultaneously on preference and personality measures. The analysis reveals that the two concepts are rather complementary when it comes to explaining heterogeneity in important life outcomes and behavior.

In Chapter 2 we seek to extent the framework of individual heterogeneities analyzed in Chapter 1 by integrating "non-cognitive skills".[3] Although research on non-cognitive skills has recently become very popular in applied economic research, there is little agreement on what is actually meant by this concept. In labor economics e.g. non-cognitive skills are usually seen as the only broadly defined second dimension in 2-factor models (next to the cognitive component), while in behavioral economics it is seen as a superordinate concept summarizing various specific concepts which include economic preferences as well as personality measures. To contribute to a joint understanding of non-cognitive skills we relate various prototypical one-dimensional non-cognitive factors to each other and decompose them into combinations of underlying personality traits and economic preference parameters. Hereby, we shed light on what previous papers measured when using different identification strategies for the non-cognitive factor. Finally, in predicting educational success, we compare different 2-factor models (including non-cognitive and cognitive components) to a "preferred" model, which uses the personality traits and economic preferences and IQ directly. We find that the inputs used to estimate 2-factor system greatly influence what is actually measured and which conclusions are reached about the role of non-cognitive skills. The results suggests a more careful interpretation of non-cognitive skills is needed when debating their importance in determining life outcomes.

Although heterogeneities in preferences and personality play such a crucial role in determining important life outcomes such as health or labor market and educational success, little is known about their origins and determinants. To provide insights into

[2]This chapter is based on Becker et al. (2012) and is joint work with Anke Becker, Thomas Deckers, Thomas Dohmen and Armin Falk.

[3]This chapter was developed jointly with John Eric Humphries.

the process of preference formation, Chapter 3 and 4 analyze the role of the social environment and focus on the (early) childhood as a critical and sensitive period in the human development process.

To contribute to an understanding of the preference formation process, Chapter 3 proceeds in two steps.[4] In a first step we present evidence showing that breastfeeding duration is a valid measure of quality of early life circumstances. In the main analysis, we secondly investigate how early life circumstances affect the formation of fundamental economic preferences such as time, risk and social preferences. In a sample of preschool children we find that longer breastfeeding duration is associated with higher levels of patience and altruism as well as a lower willingness to take risk. We repeat the analysis on a sample of young adults, which allows us to test whether the observed pattern is enduring and persists into adulthood. The results exactly mirror those found in preschool children. Importantly, in both data sets the pattern is robust to controlling for cognitive ability and socio-economic family environment. Moreover, we report evidence on health-related behavior and outcomes that are predicted based on the relation between breastfeeding duration and preferences. Finally, using data from a representative panel, we find the same preference pattern arising in response to historical variations in breastfeeding duration on a cohort level. Altogether, our findings strongly suggest that early life circumstances as measured by breastfeeding duration systematically and persistently affect human preference formation.

In Chapter 4 we build on the findings presented in Chapter 3 and, with a focus on prosociality, provide a straightforward analysis of the causal role of social environment.[5] Prosociality pervades human societies, is of fundamental importance at all levels of social interaction and contributes to economic, political and social success. Therefore it is an essential question for the well-being of individuals and societies how humans acquire prosocial attitudes. Here we present descriptive and causal evidence on the role of social environment for the formation of prosociality, measured in terms of altruism, trust and other-regarding behavior. In a first step we

[4] This chapter is based on joint work with Armin Falk.

[5] This chapter was developed in collaboration with Thomas Deckers, Armin Falk and Hannah Schildberg-Hörisch.

provide descriptive evidence on parental background and show that socio-economic status (SES) as well as mothers' prosocial attitudes systematically affect primary school children's prosociality. Children from a low SES parental background show lower levels of prosociality than children from a high SES background. Moreover, we find a positive and significant association between the prosociality of mothers and their children. This sets the stage for studying the causal role of investments in low SES children. We present evidence on a randomly assigned variation in life-circumstances, providing children with a mentor for the duration of one year. Our data reveal a significant increase in altruism, trust and other-regarding behavior in the treatment relative to the control group. These findings thus provide evidence in favor of a causal effect of social environment for the formation of prosociality. Our data additionally reveal that the investment under study substantially reduces the observed developmental gap in prosociality between low and high SES children. Finally we show that investments are most effective for children whose mothers score relatively low on our prosociality measure. In combination with the fact that mentors are particularly prosocial, this suggests that the mentoring program serves as a substitute for prosocial stimuli at the household level.

Chapter 5 takes another perspective and directly analyses the causal role of social environment on life outcomes.[6] We provide a complementary approach by combining lab and field data and focus on the effect of treatment at the workplace on health outcomes. In particular we investigate physiological responses to perceptions of unfair pay. We use an integrated approach exploiting complementarities between controlled lab and representative field data. In a simple principal-agent experiment agents produce revenue by working on a tedious task. Principals decide how this revenue is allocated between themselves and their agents. Throughout the experiment we record agents' heart rate variability, which is an indicator of stress-related impaired cardiac autonomic control and has been shown to predict coronary heart diseases in the long-run. Using three measures of perceived unfairness our findings establish a link between unfair payment and heart rate variability. Building on these findings, we further test for potential adverse health effects of unfair pay using data

[6]This chapter was developed in collaboration with Armin Falk, Ingo Menrath, Johannes Siegrist and Pablo Emilio Verde.

from a large representative data set. The analysis includes cross-sectional and dynamic panel estimations. Complementary to our experimental findings we find a strong and highly significant negative association between health outcomes, in particular cardiovascular health, and the perception of unfair pay.

Chapter 1

The relationship between economic preferences and psychological personality measures

1.1 Introduction

Both economists and personality psychologists seek to identify determinants of heterogeneity in behavior. Economists typically depict decision problems in a framework of utility maximization. An individual's utility is shaped by preferences such as risk preferences, time preferences and social preferences.[1] These preferences in combination with expectations of future events, perceptions, beliefs, strategic consideration, prices and constraints shape behavior. Personality psychology, the branch of psychology studying personality and individual differences, offers several frameworks

[1] In the standard expected utility framework, risk preference is captured by the curvature of the utility function, while the degree of risk aversion is represented in the concavity of the utility function (e.g. Gollier, 2004). Time preference describes how an individual trades off utility at different points in time (Samuelson, 1937; Frederick et al., 2002). Social preferences capture the idea that an individual's utility does not only depend on his or her own material payoff, but that it is also shaped by others' behavior and material payoff. Social preferences include altruism (e.g. Eckel and Grossman, 1996) and negative and positive reciprocity (e.g. Falk and Fischbacher, 2006). Finally, trust describes an individual's belief about others' trustworthiness combined with a preference to take social risks (e.g. Fehr, 2009). Another important economic preference is the preference for work vs. leisure. This preference is difficult to measure in experiments and is therefore not part of our analysis.

describing universal traits and individual differences. Personality traits – defined by Roberts (2009, p. 140) as “the relatively enduring patterns of thoughts, feelings, and behaviors that reflect the tendency to respond in certain ways under certain circumstances” – are an important determinant of personality (Roberts, 2006), and affect outcomes. There has been a long tradition in personality psychology to measure personality traits. The Big Five or five-factor model is the most widely used taxonomy of personality traits. It originates from the lexical hypothesis of Allport and Odbert (1936) which postulates that individual differences are encoded in language (see Borghans et al. 2008). After years of research in this tradition, psychologists have arrived at a hierarchical organization of personality traits with five traits at the highest level. These Big Five traits, that are commonly labeled as openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism, capture personality traits at the broadest level of abstraction. Each of the Big Five traits condense several distinct and more narrowly defined traits. It has been argued that the bulk of items that personality psychologists have used to measure personality can be mapped into the Big Five taxonomy (see, e.g., Costa and McCrae, 1992).[2] Another important concept in psychology focusing on individual beliefs and perceptions is the Locus of Control framework by Rotter (1966). It represents the framework of social learning theory of personality and refers to the extent people believe they have control over events.

An integration of the different measures and concepts used by economists and personality psychologists promises much potential for amalgamating evidence about the drivers of human behavior which has been accumulated disjointedly in the fields of economics and psychology (Borghans et al., 2008). Recently, scholars have begun to integrate personality into economic decision making (e.g., Borghans et al. 2008). Almlund et al. (2011) enrich theory by incorporating personality traits in a standard economic framework of production, choice, and information. Their model interprets measured personality as a “construct derived from an economic model of preferences, constraints, and information” (Almlund et al., 2011, p. 3). However, empirical knowledge is too limited to judge how personality traits relate to the con-

[2]For a more detailed description of the research on the development of the Big Five, criticism of the approach and alternative measurement systems see Borghans et al. (2008).

cepts and parameters economists typically model to predict behavior.

To shed more light on the relationship between economic preferences and psychological measures of personality we therefore study how key economic preferences, such as risk preferences, time preferences or social preferences, are linked to conventional measures of personality, such as the Big Five and Locus of Control. We analyze this relationship in a coherent framework using two main approaches. Our first approach focuses on assessing the magnitude of the correlations between psychological and economic measurement systems in three unique data sets. Our second approach departs from the fact that both preference measures and measures of personality traits predict a wide range of important life outcomes. If these two measurement systems are closely linked, they are expected to be substitutes in explaining heterogeneity in behavior. If, however, preferences and personality traits capture different aspects of behavior the two measurement systems may have complementary predictive power for important life outcomes. We therefore evaluate the individual as well as the joint explanatory power of economic preferences and psychological measures of personality in explaining health, educational and labor market outcomes.

We use three complementary datasets. First, we look at data from laboratory experiments. Using a student subject pool we conducted choice experiments on key economic preferences, namely risk taking, time discounting, altruism, trust, positive and negative reciprocity. We incentivized decision-making and obtained multiple behavioral measures for each preference. We assessed the Big-Five domains using the 60-item NEO-FFI (Costa and McCrae, 1989) and a 15 item subset, the so called BFI-S (Gerlitz and Schupp, 2005). We also measured Locus of Control using ten items adapted from Rotter (1966). Our second data set comprises very similar incentivized experimental measures with respect to risk taking and time discounting using a representative sample of almost 1000 participants from the German population. We are therefore able to obtain incentivized preference measures for a representative population. Personality was assessed using the BFI-S. The third data set stems from the German Socio-Economic Panel Study (SOEP), comprising preference and personality measures for a representative sample of more than 14.000 individuals. Preference measures were obtained using subjective self-assessment survey items rather than incentivized experiments, and personality was measured by using the BFI-S and the

Locus of Control questionnaire. Using this data set we analyze associations between important life outcomes, such as labor market success, subjective health status or life satisfaction, and individuals' preferences and personalities.

These three data sets allow for a comprehensive analysis. The first data set contains very detailed personality measures in combination with multiple experimental indicators for preferences. This student sample therefore provides a particularly accurate assessment of potential relations between economic preferences and personality. The second data set uses experimental measures for a limited set of preferences and a shorter version of the Big Five but a representative sample. Comparing results of the two data sets therefore informs us about the generalizability of our findings from the student sample. The third data set additionally allows us to study an even larger sample and to explore the explanatory power of personality and preferences for important life outcomes.

We start by analyzing data on 489 university students. We relate all five factors that capture personality according to the Big Five taxonomy and the measure of Locus of Control to our experimental preference measures. We generally find only small correlations between personality traits and preferences. In particular, only 11 of the 36 correlations in our student sample exceed 0.1 in absolute value and only one correlation exceeds 0.2 in absolute value. These eleven correlation coefficients are all significant at conventional levels, and eight of them involve correlations between social preferences and personality traits.

Next, we gauge whether the correlation patterns generalize to representative samples. We first turn to the data set that contains very similar experimental measures of risk and time preferences and survey measures of the Big Five for about 1000 individuals, who were sampled to be representative of the adult population living in Germany (see Dohmen et al., 2010). The correlation structure between personality traits and risk and time preferences turns out to be similar to the one we find for students, with few exceptions.

Finally, we assess whether the empirical associations between preference parameters and personality traits are sensitive to the way in which preferences are measured. We compare correlations between personality traits and measures of preferences derived from the incentivized choice experiments in the student and the representative sample to correlations that are constructed based on the non-incentivized subjective self-assessments in a representative sample of 14.000 individuals from the SOEP. Our result on the pattern of correlations between preference measures and personality measures is again largely confirmed.

We then turn to a different type of analysis in which we assess the power of preferences and personality in explaining life outcomes, including health, life satisfaction, earnings, unemployment and education. Our analysis reveals that both measurement systems have similar explanatory power when used separately as explanatory variables. The explained fraction of variance increases by about 60% when life outcomes are regressed on both measurement systems. We therefore conclude that each measurement system captures distinct sources of the heterogeneity in life outcomes. A coherent picture emerges from our analysis. Both approaches strongly suggest that standard measures of preferences and personality are complementary constructs.

So far no clear picture concerning the relations between measures of personality and economic preferences has emerged in the literature (see Almlund et al., 2011). For example, the study by Daly et al. (2009) suggests a negative relationship between conscientiousness and the discount rate, but such a negative correlation is neither corroborated by Dohmen et al. (2010), who relate experimental measures of willingness to take risk and impatience to survey measures of the Big Five in a representative sample of adults living in Germany, nor by Rustichini et al. (2012), who relate a measure of delay acceptance to four of the Big Five domains in a sample of 1065 U.S. trainee truckers.[3] In fact, Dohmen et al. (2010) find no significant relationship between personality traits and preference measures in a regression framework that includes controls for IQ, gender, age, height, education, and household income. Raw correlations between preference and personality measures, which are also reported in Almlund et al. (2011), are weak; time preference is significantly correlated to agreeableness only (at the 10 percent level).[4] This finding is confirmed by the significant correlation between delay acceptance and agreeableness in the truck driver

[3] The effect sizes of the correlations between preference and personality measures are all smaller than 0.1 in absolute value.

[4] We report this data in Table 1.3.

sample of Rustichini et al. (2012).

Evidence on the link between risk preferences and Big Five domains is equally mixed. Raw correlations between a lottery choice measure of risk preference and personality traits in the data from Dohmen et al. (2010) indicate significant relationships between risk preferences and openness to experience (at the 1 percent level) and with agreeableness (at the 5 percent level). Rustichini et al. (2012) do not measure openness to experience. They do not find a significant correlation for risk preference and agreeableness, but report a weak correlation between risk preference and neuroticism (0.05 in absolute value), which is significant at the 10 percent level. This finding is in line with the significant positive association between risk aversion and neuroticism reported by Borghans et al. (2009). Other researchers (e.g. Zuckerman, 1994) have related risk preferences to sensation seeking, a facet of extraversion in the Big Five taxonomy, and found mixed evidence. While Bibby and Ferguson (2011) report a significant correlation between a measure of loss aversion and sensation seeking ($r = 0.27$), Eckel and Grossman (2002) find no evidence of an association between risk preferences and sensation seeking.

Evidence on the link between social preferences and personality is somewhat stronger. Dohmen et al. (2008) relate survey measures of social preferences to measures of the Big Five using data from the German Socio-Economic Panel Study (SOEP) and find significant associations between trust, as well as positive and negative reciprocity and personality traits. Trust is positively related to agreeableness and openness to experience, and negatively to conscientiousness and neuroticism; while positive reciprocity is positively associated with all five personality factors, negative reciprocity is negatively related to conscientiousness and extraversion, and positively to neuroticism. A link between extraversion and behavior in the dictator game, which can be interpreted as a measure of altruism, has been established by Ben-Ner and Kramer (2011).

The paper is structured as follows. Section 1.2 describes our three data sets. In section 1.3 we introduce our research strategy for investigating the link between personality and preferences. Section 1.4 presents evidence on the correlation between measures of personality and measures of preferences. In addition it contains an assessment of the explanatory power of preferences and personality in explaining important life outcomes. Section 1.5 concludes.

1.2 Data and measures

In this section we provide a description of the three complementary data sets that we employ for our analysis. Before we present our experimental and survey measures in detail a few comments on identification are warranted. Economists typically try to infer preferences from choices, the so-called revealed preference approach. For example, one might surmise that a person who does not wear a safety belt or does not invest in risky stocks has a preference for taking risks. It is, however, easy to show that the same behavioral pattern is compatible with very different risk preferences if other factors affect the person's decisions. For example, differences in beliefs about how risky driving without a safety-belt or investing in stocks actually is may affect decisions equally strong than underlying risk preferences. The problem is that the decision context is uncontrolled and person specific, rendering precise statements about preference parameters very difficult.[5] This is why economists run experiments to infer preferences. In a typical choice experiment subjects take decisions in a well controlled decision environment. In risk experiments, e.g., stakes and probabilities are fixed and the action space is identical for every subject. Observing subjects' decisions in a controlled experimental environment therefore rules out many potentially confounding factors, allowing a more precise identification of preferences. Even in an experiment, however, identification of preferences is limited (see Manski (2002) for a thorough discussion on identification of experimental outcomes). The same observed action can reflect different risk attitudes, e.g., if the experimental subjects dispose of different wealth levels and the curvature of the utility function is not invariant to wealth levels. Despite these limitations experiments deliver much more precise behavioral outcomes than non-experimental observations. In strategic situations, which are relevant for measuring trust and reciprocity, we are able to

[5]Conceptually identical problems apply to the identification of traits, such as ability, physical strength and personality characteristics from observed performance on tasks, when performance also depends on other unobserved factors such as time, energy or attention devoted to the task. An illuminating discussion of the identification problem is provided in section 1.3 of Almlund et al. (2011).

elicit not just an action but a complete strategy. With field observations this is impossible. The relevance of eliciting a strategy is obvious: Suppose observing a second mover who defects in a cooperation context, in response to a non-cooperative act of a first mover. This could reveal selfish preferences as well as reciprocal preferences. To disentangle the two requires knowledge about what the decision maker would have done, had the first mover cooperated. Eliciting a strategy instead of observing only actions does exactly this. Experimental observations have the additional advantage over survey responses that decisions have immediate monetary consequences. This is of obvious importance, e.g., for identifying altruism. It makes a big difference to simply state altruistic preferences or to reveal them in a costly manner.

1.2.1 Experimental data

The first data set consists of decisions from laboratory experiments among university students. We ran a series of simple incentivized choice experiments to elicit preferences concerning risk taking, discounting, positive and negative reciprocity, trust as well as altruism.[6] Table 1.1 presents an overview of the experiments and provides a short description of the elicitation methods and the obtained behavioral measures. Four important features about our experimental design are worth noting. First, for risk taking, discounting, trust and positive reciprocity subjects took part in two very similar experiments each. This allows us to average over both outcomes for each subject in order to minimize measurement error. Second, in order to reduce spillovers between different choices, experiments were not run in one single session but in two sessions, which were scheduled one week apart from each other.[7] Third, in order to reduce possible income effects with respect to outcomes within session, feedback about experimental outcomes was only given at the very end of an experimental session. Fourth, the vast majority of subjects in the experiments had never taken part in an experiment before. This eliminates possible confounds in behavior due to previous experiences in similar experiments. In total, 489 students from different

[6]For a detailed description of the experimental procedures see Falk et al. (2011).

[7]We reversed the order of the sessions for half of the subjects. Statistical tests reveal no significant order effects.

majors from the University of Bonn took part.[8] The experiments were run at the Laboratory for Experimental Economics at the University of Bonn (BonnEconLab). We used zTree (Fischbacher, 2007) as experimental software and recruited subjects using ORSEE (Greiner, 2004). Each session lasted about two hours, and average earnings were 64 Euros.

Table 1.1: Overview: Experimental measures

Preference	**Experiment**	**Measure**
Time	Two lists of choices between an amount of money "today" and an amount of money "in 12 months".	Average switching point over both lists of choices from the early to the delayed amount.
Risk	Two lists of choices between a lottery and varying safe options.	Average switching point over both lists of choices from the lottery to the safe option.
Positive Reciprocity	Second mover behavior in two versions of the Trust Game (Strategy Method).	Average amount sent back in both Trust Games.
Negative Reciprocity	Investment into punishment after unilateral defection of the opponent in a Prisoner's Dilemma (Strategy Method).	Amount invested into punishment.
Trust	First mover behavior in two versions of the Trust Game.	Average amount sent as a first mover in both Trust Games.
Altruism	First mover behavior in a Dictator Game with a charitable organization as recipient.	Size of donation.

1.2.1.1 Preference measures

Risk preferences In order to elicit risk attitudes we adapted the design from Dohmen et al. (2010). Subjects were shown a list of binary alternatives, a lottery

[8]Out of these 489 students, 80 took part in a pretest of the study. Most of these 80 subjects had taken part in an experiment before. The pretest did not include the experiments on altruism and negative reciprocity.

and a (varying) safe option. The lottery was the same for each decision: if they chose the lottery participants could either win 1000 points or zero points with 50 percent probability each. The safe option increased from row to row, starting from a value of (close to) zero, and increasing up to a value of (close to) the maximum payoff of the lottery. To reduce measurement error subjects participated in two risk experiments. The choice list of the second experiment was simply a perturbed version of the first one. Perturbations were constructed such that a randomly drawn integer value between -5 and +5 was added to the safe option in every choice, corresponding to perturbations of maximally 5% of the step size of the increase in the safe option. The complete list of choices was shown to subjects on the first screen. Each choice situation was then presented on a separate screen, where subjects entered their respective choice. Subjects were informed that one choice in each list would be selected randomly and paid. Subjects with monotonic preferences should choose the lottery for lower safe options and switch to the safe option when the latter reaches or exceeds the level of their certainty equivalent. Thus, switching points inform us about individual risk attitudes. The earlier a subject switches to the save option the less he or she is willing to take risks. For our analysis we constructed a risk preference measure using the average of the two switching points from the two experiments.[9]

Time preferences To measure individuals' time preferences we implemented a procedure very similar to the one for risk attitudes. In the discounting experiments, subjects were given two lists of choices between an earlier amount of money ("today"), which was the same in all choices, and an increasing delayed amount of money ("in 12 months"). In the first row the early amount was equal to the delayed amount. Delayed amounts increased from row to row by 2.5%. As for risk preferences subjects participated in a very similar second discounting experiment with small perturbations of delayed amounts between +0.5 and -0.5 percentage points. One choice in each of the two lists was randomly selected for payment. Payments resulting from the two experiments were sent to subjects via regular mail. If a subject chose the

[9]If subjects switched between the lottery and the safe option more than once, we took the average switching row as an estimate of their certainty equivalent. This happened in 16 % of the cases in the first experiment on risk taking, and in 11 % of the cases in the second experiment.

early amount, the payment was sent out on the day of the experimental session. If a subject chose the delayed amount, the payment was sent out with a delay of 12 months.[10] The switching point from early to delayed payment informs us about a subject's time preference. Subjects who switch later discount the future amount by more, i.e., are less patient, than subjects who switch earlier.[11] Our measure of individual discounting is the average switching row in both lists. To ease interpretation of the correlations reported below, we recode the measure, such that higher values imply earlier switching rows, i.e., a higher level of patience.

Trust We elicited trust from first mover behavior in the so-called Trust Game (Berg et al., 1995). We conducted two versions of the Trust Game. In one version, the amount sent by the first mover was doubled by the experimenter, while in the second version the amount was tripled. Every subject was in the role of the first and of the second mover twice.[12] Both Trust Games were incentivized, i.e., every (relevant) decision was paid. In the role of a first mover subjects could choose to send any amount in $\{0, 50, 100, \ldots, 500\}$ points to the second mover. All interactions in the Trust Game as well as in all other social preference experiments were one-shot and anonymous (perfect stranger matching protocol). The average amount sent as a first mover in both Trust Games constitutes our experimental measure for trust: subjects who send higher amounts of money are those who display higher levels of trust.

Positive reciprocity To elicit positive reciprocal inclinations we measure subjects' second mover behavior in the Trust Game (compare previous paragraph). We implemented the Strategy Method (Selten, 1967). This means that for every possible amount sent by the first mover, subjects were asked to indicate how much they wanted to send back. The actual decision of the first mover determined which of these decisions became payoff relevant. The average amount sent back as a second

[10]Keeping the payoff mode identical over both time horizons rules out credibility concerns.

[11]For subjects, who switched more than once, we took the average switching row as an estimate of their discount rate. This happened in 5 % of the cases in the first experiment on time discounting, and in 7 % of the cases in the second experiment.

[12]Overall, we therefore ran four Trust Games.

mover in both Trust Games was taken as individuals' willingness to reciprocate, such that higher values imply a higher willingness to reciprocate.

Negative reciprocity In order to measure subjects' willingness to engage in costly punishment of unfair behavior, we conducted a Prisoner's Dilemma with a subsequent punishment stage.[13] In the punishment stage, subjects could choose to invest points in order to deduct points from their opponent. Punishment was costly. Again, we implemented the Strategy Method. Before taking their decisions in the first stage of the experiment, i.e., in the Prisoner's Dilemma, subjects were asked to indicate how many points they wanted to deduct from the other player in case he or she cooperated or defected, for both own cooperation and own defection. Then, they played a simultaneous Prisoner's Dilemma. The outcome of the first stage determined which choice of the second stage became payoff relevant. The chosen investment into punishment after unilateral defection of the other player served as a measure of an individual's willingness to reciprocate negatively.

Altruism To measure altruistic behavior subjects took part in a modified Dictator Game in which the recipient was a charitable organization (adapted from Eckel and Grossman, 1996). Subjects were endowed with 300 points and had to decide how much of this endowment to donate to a charitable organization.[14] This decision serves as our experimental measure of subjects' altruistic inclination.

1.2.1.2 Personality measures

Big Five As part of the study, subjects were given a paper-and-pencil survey, which they were asked to fill out at home and return to us via mail.[15] 319 out of 489 subjects completed the survey and sent it back to us. The survey included the NEO-FFI version of the Big Five (Costa and McCrae, 1989). During the experimental sessions all 489 subjects also answered a shorter version of the Big Five, consisting

[13]The design of the experiment was adapted from Falk et al. (2005)

[14]Subjects could choose a charitable organization from a list, or name one themselves.

[15]We also handed out stamped envelopes with the address of our research institute, in order to minimize additional costs for returning the survey to us.

of 15 items which are a subset of the NEO-FFI. This so-called BFI-S has been developed by Gerlitz and Schupp (2005) and was also part of the 2005 and 2009 waves of the German Socio Economic Panel (SOEP). Correlations between the long version and the short version of the Big Five differ between the five personality dimensions. The lowest correlation is $r = 0.48$ for openness, the highest is $r = 0.71$ for conscientiousness, respectively (all p-values < 0.001). We constructed our Big Five measure in that we use data from the long version whenever available, while for the remaining subjects we refer to the short version. That way, we have measures of all Big Five domains for all 489 subjects.

Locus of Control The paper-and-pencil survey included 10 items that allows us to construct a measure of Locus of Control for the 319 individuals who filled in the survey. These 10 items have been adapted from Rotter (1966) and they have also been implemented in the 2005 wave of the SOEP. The personality construct of Locus of Control assesses in how far a person believes to have control over their life outcomes, or in how far their life is determined by forces that are outside of their control, such as luck or faith. We constructed the measure such that higher values represent a more internal Locus of Control, i.e., the belief that the person can influence their life outcomes. Lower values represent a more external Locus of Control.

1.2.2 Representative experimental data

The second data set we employ consists of experimental data for a representative sample of the German population.[16] This data set is used to assess whether the findings from the sample of university students can be corroborated in a representative sample. Subjects' risk and time preferences were elicited, and we again have information on participants' personality. The data used here stem from a study conducted in 2005 and contains information on 1012 individuals. For a detailed description of the study and its procedures see Dohmen et al. (2010).

[16]The same data set is used in Dohmen et al. (2010).

Preference measures The experiments on risk and time preferences were similar to the ones we used in the laboratory experiments. In both experiments subjects had to make multiple decisions in a list of choices. To elicit their risk preferences subjects chose between a lottery, which remained the same in all choices, and safe options, which increased in their value. Like before, the switching point informs us on the individual's willingness to take risks. similarly, to elicit individuals' time preferences all participants made a number of intertemporal choices. They had to decide between an amount "today" and a larger amount 12 months later. The early amount remained the same in all choices. The first delayed amount presented to subjects was devised to imply a 2.5% return on the early amount assuming semi-annual compounding. In the subsequent choices the delayed payment was gradually increased and was calculated such that the implied rate of return rose in steps of 2.5 percentage points. As before, the switching points from the early to the delayed option inform us on the subjects' time preferences.

Personality measures The five personality domains were assessed using the BFI-S (see section 1.2.1.2 for a more detailed description).

1.2.3 Representative panel data

The third data set we use stems from the German Socio-Economic Panel Study (SOEP), a large panel data set that is representative of the adult population living in Germany (see Schupp and Wagner (2002) and Wagner et al. (2007) for a detailed description of the SOEP). We use information from eight waves collected in the years between 2003 and 2009. In each of these waves more than 20,000 individuals were interviewed. The SOEP combines extensive socio-demographic information with various measures of attitudes, preferences and psychological traits. In particular, the SOEP includes survey items relating to all personality and preference measures that we have analyzed in the previous sections.

Personality and economic preference measures were elicited several times between 2003 and 2009. To construct a measure for each individual, we use the maximum available number of observations of a given measure. If several measures of personality and preferences are available, we take the average of the standardized measures of all years in which this measure was elicited. The resulting average is then standardized as well. In case a particular measure was elicited only in one wave (as it is the case for patience, for example) we just take the standardized measure from that respective year. We restrict the sample to individuals for whom we have information about each personality and preference measure. This results in a sample size of 14,243 individuals.

Preference measures As a measure for time preference we use answers to the following survey question: "How would you describe yourself: Are you generally an impatient person, or someone who always shows great patience?".[17] Participants gave an answer on an 11-point scale where zero means "very impatient" and ten means "very patient". This item was implemented only in 2008. The risk preference question was asked in the same manner: "How do you see yourself: Are you generally a person who is fully prepared to take risks, or do you try to avoid taking risks?" Answers were given on an 11-point scale where zero means "unwilling to take risks" and ten means "fully prepared to take risks". This question was asked in the four waves 2004, 2006, 2008 and 2009. The general risk question has been studied in various papers and has been validated using incentivized experiments in representative samples as well as using behavioral evidence in Dohmen et al. (2011). In 2005 the SOEP contained six items to measure reciprocal inclinations, three items each on positive and negative reciprocity. Examples for positive and negative reciprocity are: "If someone does me a favor, I am prepared to return it" and "If I suffer a serious wrong, I will take revenge as soon as possible, no matter what the costs". Participants expressed how well these six statements apply to them on a 7-point Likert scale. For a detailed description see Dohmen et al. (2009). Standard trust questions were asked in the two waves 2003 and 2008, using three sub-statements about whether "one can trust people", whether "in these times one can't rely on anybody else" and whether "when dealing with strangers it is better to be cautious". Answers were given on a 5-point scale ranging from "Totally agree" to "Totally disagree". Finally, our survey measure

[17]The behavioral validity of this question with respect to incentivized experiments is documented in Vischer et al. (2013).

for altruism is the answer to the question how important it is for the participant "to be there for others". Answers were given on a 4-point scale. The altruism question was asked in waves 2004 and 2008.

Personality measures The 2005 and the 2009 wave of the SOEP contained the BFI-S questionnaire, developed by Gerlitz and Schupp (2005). Locus of Control was elicited in 2005 using Rotter's Locus of Control scale (Rotter, 1966). Both inventories were also used in our lab experimental data. See section 1.2.1 for more details on the BFI-S and the Locus of Control scale.

1.3 Research strategy

To answer the question whether measures of personality and economic preferences are closely linked we first study the raw correlations between these measures. High correlations would indicate some degree of substitutability. Low correlations, on the other hand, would suggest that the two measurement systems are complementary concepts in explaining heterogeneity in behavior. Whether a correlation should be interpreted as "high" or "low" is of course always debatable. We therefore first look at statistical significance levels. Statistical significance, however, can also be found for correlations which are low in terms of effect size (Cohen, 1992). Following conventions in the social sciences we interpret effect sizes, i.e., correlations r, as rather "low" if r is between 0.1 and 0.3, as "medium" if r is between 0.3 and 0.5 and as "large" if r is larger than 0.5. Since the analysis of correlations is restricted to linear relations, we also check for potential non-linear associations by conducting non-parametric regressions. In particular, we look at Kernel-weighted local linear polynomial regressions.

We then check whether measures of personality and preferences are substitutes or complements in terms of their explanatory power for life outcomes. In particular, we conduct linear regressions and assess the explanatory power of the two concepts by reporting levels of adjusted R-squared. In these regressions, measures of personality and preferences are included individually as well as jointly. If the two measurement systems are substitutes, adjusted R-squared in the combined regressions should not be distinctly higher than in regressions in which only one of the two concepts is included. The opposite should hold for complements. Additionally, we investigate model selection criteria in these regressions. We check for robustness using binary and ordered choice models as well as more comprehensive specifications including square terms and cross-products of all regressors.

1.4 Results

In this section we discuss our main findings. Note that in order to ease comparison between data sets and measures all experimental as well as all personality measures were standardized for the data analysis.

1.4.1 Correlation structure

1.4.1.1 Experimental data

Table 1.2 displays the 36 raw correlations of the personality and economic preference measures obtained from the lab experiments. A first inspection of Table 1.2 reveals that only eleven of these 36 correlations are statistically significant at the 5% or 1% significance level.[18] All correlation coefficients are smaller than 0.3 in absolute value. Hence, there is no correlation with "medium" effect size or larger. Moreover, of all 36 correlations only eleven exceed 0.1 in absolute value and only one of these slightly exceeds 0.2.[19]

Table 1.2 also shows that among all personality factors agreeableness exhibits the highest and statistically most significant correlations with measures of economic preferences. It is significantly correlated with measures for positive and negative reciprocity, trust and altruism (all p-values < 0.01) as well as with time preference

[18]Five additional correlations are weakly significant, i.e., significant at the 10% significance level.

[19]Results qualitatively stay the same when investigating Spearman correlations instead of Pearson correlations (see Table A1.2 in Appendix A1). Moreover, when looking at a potential linear mapping, i.e., linear regressions of either the Big Five on preferences or vice versa, R^2 is always below 10%.

Table 1.2: Correlation structure experimental data set

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism	LoC
Time	0.0370	0.0057	−0.0084	0.1026**	−0.0518	0.0847
Risk	−0.0379	−0.0611	0.0762*	0.0202	−0.1201***	0.0434
Pos. Reciprocity	0.1724***	0.0140	0.0211	0.2042***	0.0361	0.0152
Neg. Reciprocity	−0.0885*	−0.0393	0.0943*	−0.1451***	−0.0136	−0.1418**
Trust	0.1232***	−0.1300***	0.0004	0.1665***	−0.0134	−0.0140
Altruism	0.1242**	−0.0979*	0.0249	0.1911***	0.0847*	0.0480

Pearson correlation coefficients. *, **, and *** indicate significance at the 10%, 5%, and 1% level. Correlations between economic preferences and the Big Five were calculated using between 394 and 477 observations. Correlations between economic preferences and Locus of Control were calculated using between 254 and 315 observations. All measures are standardized.

(p-value < 0.05). Correlations with social preferences are all in the range between 0.1 and 0.3 in absolute value, indicating a small effect size according to the classification of Cohen (1988). The high frequency of significant correlations of agreeableness with social preferences is not surprising as the former is defined as "the tendency to act in a cooperative, unselfish manner, ... " (see Table A1.1).

Finding only moderate correlations between preference and personality measures does not necessarily indicate that these constructs are weakly connected; it only indicates that there are weak linear relations. For example, a perfect U-shaped relation between a personality factor and a preference would result in an insignificant linear correlation. To explore the possibility of non-linear relationships we therefore estimate Kernel-weighted local linear polynomial regressions.[20] In each regression, we restrict the sample to a range of four standard deviations around the mean of each variable to circumvent an analysis biased by outliers. Therefore, the results are calculated using 70% to 97% of all observations. The predicted regressions are displayed in Figure A1.1. Although sometime there are small deviations from linearity at the boundaries, the overall picture strongly suggests a linear relation in the vast majority of combinations.

Summarizing our analysis of the lab experimental data we find that associations between preference and personality measures are linear and that the degree of association is rather low, suggesting a complementary relationship. We next turn to the question whether the correlation patterns observed in student samples can be replicated in a sample that is representative of the adult population.

[20]We use the Epanechikov kernel and bandwidth is selected via the plugin estimator of the asymptotically optimal constant bandwidth.

1.4.1.2 Representative experimental data

Table 1.3: Correlation structure representative experimental data

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Time	−0.0080	−0.0682	−0.0655	−0.0830*	−0.0602
Risk	0.1356***	−0.0720	0.0757	−0.0941**	−0.0290

Pearson correlation coefficients. *, **, and *** indicate significance at the 10%, 5%, and 1% level. All measures are standardized.

Table 1.3 shows the correlations between the outcomes from the risk and time experiments and the personality traits. As before, the measure for time is reversed, so that higher values indicate higher patience. In terms of significance the pattern is similar to the one in the laboratory study. Only one correlation is significant at the 1%-level, one is significant at the 5%-level and one is significant at the 10%-level. In terms of effect size, only the coefficient of the association between openness and risk preferences exceeds the 0.1 benchmark to be classified as a small correlation (Cohen, 1988).[21] Interestingly, the sign is positive, in contrast to our laboratory data. The other two significant coefficients are even smaller. The analysis of representative data therefore confirms that the level of association between preference personality measures is rather small. However, we can only draw this conclusion with respect to time and risk preferences, as we do not have experimental data on trust and social preferences. We next analyze whether these findings also hold when looking at all preference measures in a large representative sample.

1.4.1.3 Representative panel data

In this section we study whether our findings from the experiments generalize to a large representative sample using survey rather than experimental instruments for measuring economic preferences. Table 1.4 shows the raw correlations between

[21] Results qualitatively stay the same when investigating Spearman correlations instead of Pearson correlations (see Table A1.3 in Appendix A1).

Table 1.4: Correlation structure SOEP

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism	LoC
Time	0.0183**	0.1122***	−0.0415***	0.3122***	−0.0584***	0.0681***
Risk	0.2793***	−0.0400***	0.2601***	−0.1454***	−0.0996***	0.1521***
Pos. Reciprocity	0.1814***	0.2520***	0.1473***	0.1842***	0.0872***	0.0954***
Neg. Reciprocity	−0.0522***	−0.1558***	−0.0264***	−0.3756***	0.0612***	−0.2154***
Trust	0.1272***	−0.0680***	0.0575***	0.0945***	−0.1919***	0.2094***
Altruism	0.1756***	0.1495***	0.1670***	0.2557***	0.0908***	0.0874***

Pearson correlation coefficients. *, **, and *** indicate significance at the 10%, 5%, and 1% level. Correlations are calculated using 14,243 observations. All measures are standardized.

personality measures and economic preferences using 14,243 observations from the SOEP. Given the large number of observations it is not surprising to find a large number of significant correlation coefficients (p-values < 0.05 for all correlation coefficients). In terms of effect size, however, only two correlations are of "medium" size, i.e., larger than 0.3. 18 of the reported 36 correlations can be classified as being "small", while 16 correlations are even below 0.1. This confirms the overall picture which emerged from the analysis of the two experimental data sets.[22] A closer comparison of the SOEP survey measures with our experimental measures further reveals large similarities. As reported above, eleven correlations are significant at the 5% level in the experimental data. Ten of these correlations have the same sign and are significant at the 1% level using survey data. Moreover, as it is the case in the lab data set, it is again the personality trait agreeableness which exhibits the highest correlations with economic preferences, in particular social preferences. While there are small differences in the results compared to the experimental data set, i.e., seven out of 36 correlation coefficients show a different sign, the general pattern emerging from the SOEP measures is consistent with our previous findings. Out of the seven correlation coefficients only two are (weakly) significant in the experimental data set. Nevertheless, we think that the inconsistency of signs questions the conjecture that correlations are universally identical, i.e., identical irrespective of age or other person characteristics. We return to this aspect in the final section.

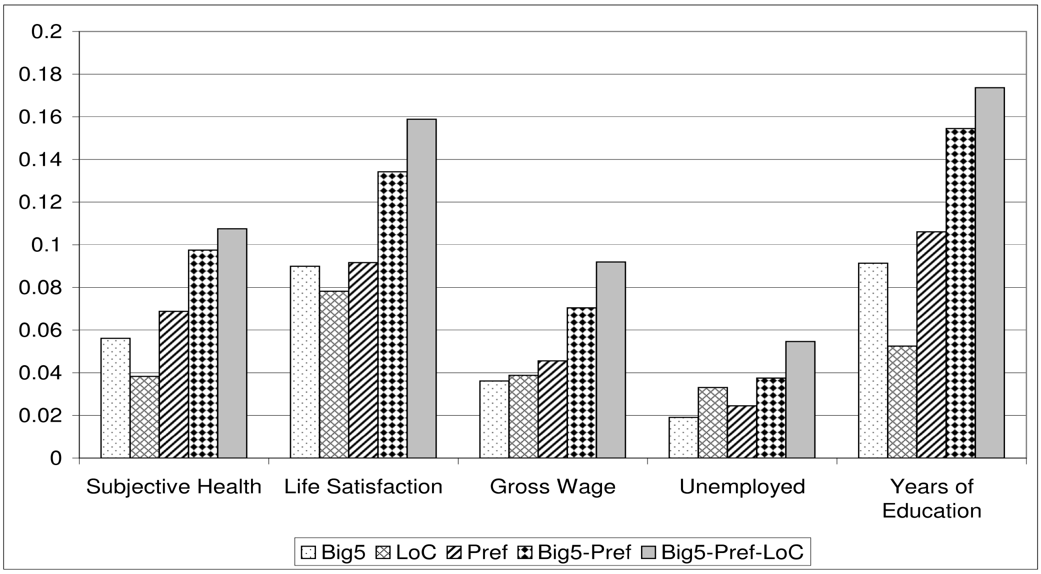
We conclude this section with an analysis of potential non-linearities between our SOEP preference and personality measures. As for the lab experimental data, we perform Kernel-weighted local linear polynomial regressions restricting the sample in each regression to four standard deviations above and below the mean. The resulting subsamples represent 92% to 97% of the observations of the main sample. The predicted functions presented in Figure A1.2 show no particular non-linearities, except for some splines at the left ends of the considered range. Thus, analogously to the experimental data set, it is not the case that systematic non-linearities bias

[22]Results qualitatively stay the same when investigating Spearman correlations instead of Pearson correlations (see Table A1.4 in Appendix A1). Moreover, when looking at a potential linear mapping, i.e., linear regressions of either the Big Five on preferences or vice versa, R^2 is always around 15% with the exception of agreeableness, where R^2 reaches 28%.

correlation coefficients.

1.4.2 Explanatory power for life outcomes

Figure 1.1: Adjusted R-squares for life outcomes



This figure shows adjusted R-squares for linear regressions. The number of observations varies for the different life outcomes: Subjective Health (14,218 obs.), Life Satisfaction (14,214 obs.), Gross Wage (7,199 obs.), Unemployed (9,095 obs.), Years of education (13,768 obs.). Gross Wage measures the gross hourly wage.

All reported correlation structures indicate that personality and preference measures are far from being perfectly substitutable. In order to determine whether they actually complement each other, we now analyze their explanatory power with respect to important life outcomes. To that end we again use data from the SOEP. In particular, we consider the following outcomes: subjective health, life satisfaction, gross wage, being unemployed and years of education. For each outcome we estimate linear regression models in which outcomes are regressed on the set of economic preferences, Big Five and Locus of Control, separately as well as jointly.[23] The idea is to assess the explanatory power of each concept in isolation and in combination. This

[23]The corresponding regressions are shown in Table A1.5 in Appendix A1.

enables us to check the extent to which explanatory power increases when combining the concepts and thus allows us to reach conclusions regarding the degree of their complementarity. The criterion used to compare differences in explanatory power is adjusted R-squared.

All life outcomes we use come from the 2009 wave of the SOEP. Subjective health was measured on a 5-point-scale, from "very good" to "bad". We reverse the answer scale such that higher values indicate a better subjective health status. Life satisfaction was elicited using the question "How satisfied are you with your life, all things considered?", which was answered on an 11-point-scale (with higher values indicating higher life satisfaction). Our measure for gross hourly wage is the gross monthly wage divided by monthly working hours.[24] Unemployment status is a binary variable equal to one if the person was unemployed at the time of the survey and zero otherwise. The variable years of education is created by adding up years of schooling and additional occupational training (including university).[25]

Figure 1.1 shows adjusted R-squares for the different life outcomes. R-squared values for the three concepts – Big Five, Locus of Control and economic preferences – in isolation are in a range between 1 and 10 percent and vary both between concepts and outcomes. Thus, they contribute to explaining heterogeneity in important life outcomes.[26] More important in light of our research question, however, is the fact that the explanatory power is considerably larger when combining Big Five, Locus of Control and economic preferences compared to using each concept individually. Moreover, explanatory power is always maximized when all three concepts are included in the regression, hereafter referred to as the full model. In this case, resulting adjusted R-squared values reach levels of about 6 to 18 percent. This clearly indicates the existence of important complementarities among the different concepts.[27]

Since the question we are answering here is a question of model selection, we also

---

[24]Monthly working hours are calculated as the average weekly working hours multiplied by four.

[25]For each school degree and occupational training (including university) official standard graduation times in years are used for the calculation.

[26]When explaining life outcomes such as gross wages, unemployment or years of education the preference for work vs. leisure would probably play a key role. However, no question relating to this preference was included in the survey.

[27]For an overview over the raw correlations between each preference and personality trait and life outcomes see Figure A1.3 and A1.4.

employ model selection criteria (in particular the Akaike and Bayesian information criterion) to check whether the full model is also chosen by model selection criteria. As can be seen in Table A1.6 in Appendix A1 this is the case for all life outcomes considered, corroborating our previous results. We perform the same analysis using binary and ordered choice models when appropriate. Again, the full model is chosen by the model selection criteria in all cases. As another robustness check we consider more flexible models: Next to including each predictor linearly in our regressions we also include square terms and all possible cross products (see Table A1.7 in Appendix A1). Again the full model obtains the highest adjusted R^2 measures when using OLS estimation and is also chosen by the information criteria in nearly all cases.[28] Results are again robust for employing binary and ordered choice models when appropriate. Moreover, in all models considered the joint hypothesis that all coefficients are equal to zero is always rejected at the 1% level (Tables A1.6 and A1.7 in Appendix A1). Summing up, sizeable complementarities among the different concepts are corroborated in all robustness checks.

1.5 Discussion

In this paper we have examined the relation between economic preferences and personality using three different data sets. We find no indication for a strong linear nor a non-linear association between the two. Thus we conclude that the two concepts cannot substitute each other. In fact, when it comes to explaining heterogeneity in life outcomes, we find that the two concepts play a complementary role. Our findings imply that researchers in economics and psychology can heavily benefit from the respective other discipline when looking for potential sources of heterogeneity in life outcomes.

Finding a rather low association between economic preferences and psychological measures of personality is perhaps not surprising. First, both concepts are constructed in very different ways. While preferences are rooted in utility theory, derived

[28]Only the BIC choses a model just including Locus of Control when it comes to explaining gross wage and unemployment. However, this is not surprising given the number of regressors included and the tendency of BIC to choose parsimonious models.

in terms of specific functional forms of utility functions, the Big Five personality indicators originate in language analysis. Second, the Big Five measure rather broad aspects of personality. In particular, each dimension of the Big Five is by itself already an aggregation of different attitudes or subfacets. Thus, while our results show low associations between personality and economic preferences, we cannot exclude the possibility that there is a stronger degree of association between economic preferences and subfacets of the five personality traits. The trait extraversion, for example, is composed of different attitudes, such as being "relatively more outgoing, gregarious, sociable, and openly expressive" (see Table A1.1), measured by 12 different questions in the NEO-FFI or three different questions in the BFI-S. Put differently, each personality measure is not only made up of multiple items, but more importantly captures distinct aspects of a character trait. Economic preferences, on the other hand, are defined more narrowly. For example, the concept of time preferences refers to the individual's willingness to abstain from something in the present in order to benefit from that decision in the future. While this concept is applicable to different domains, e.g., to health outcomes or financial decision making, the underlying concept remains the same and is measured by standard incentivized experiments or survey items as employed in this study. In this sense, our preference measures might resemble the subordinate aspects of the five personality factors.

Third, finding strong complementarities between economic preferences and personality measures may simply reflect conceptual differences in the way economic and psychological models are constructed. The economic model explains heterogeneity in behavior in terms of three distinct components: preferences, beliefs and constraints, such as abilities. In contrast, psychological measures such as the Big Five include notions of preferences as well as beliefs and constraints. In other words, in our analysis we have correlated economic preferences at least partly with beliefs and constraints, which by construction should not necessarily be correlated. A good example is conscientiousness. Being able and willing to work hard and being organized comprises aspects of both, preferences and personal abilities. Likewise, emotional instability, which is part of the neuroticism facet, is related to personal inability rather than a preference. Even more extreme is the case of Locus of Control, which is clearly a belief rather than a preference. This does not rule out the possibility that the two concepts are related, e.g. because an external Locus of Control is conducive to the development of impatient behavior: if it does not pay off to invest because life circumstances are predominantly determined by circumstances beyond my control, the willingness to forgo current consumption and wait in order to earn a return in the future makes little sense. Yet, beliefs and preferences are two distinct concepts.

The main focus of this paper is the rather weak association and complementary nature of economic and psychological measures of personality. We have not discussed the specific signs of the correlations or ways to integrate personality into the economic model. Important work in this direction has been done in Almlund et al., 2011. Many signs of the correlations reported above are consistent across the three data sets, in particular those that are significant. For example, in all three data sets risk attitudes and extraversion are positively, risk and neuroticism are negatively correlated. There are important exceptions, however. In the student sample, e.g., risk attitudes and openness are negatively correlated while they are positively and significantly negatively correlated in the two representative data sets. These and other inconsistencies raise important questions. One possible reason for finding different signs is the use of different elicitation methods for economic preferences (experiments and survey responses). Another possibility is that the reported correlations vary over the life-cycle. If traits develop with different speed and at different points in life correlations should vary with age. This could explain differences between a relatively young student sample and the representative samples. Not much is known about how economic preferences develop over the life-cycle but at least for risk attitudes there seems to be a robust and large negative age effect on willingness to take risks (Dohmen et al., 2011). Another possibility is that preferences and personality are generically differentially correlated between specific groups of the population, e.g., varying by gender, age, height or education. From an evolutionary perspective the co-evolution of traits may serve different purposes depending on specific life circumstances. It may be "optimal" for one subgroup of the population to develop a positive correlation between particular traits, while for another subgroup it is adaptive to form a negative correlation. More work needs to be done to uncover potential group specific correlations between personality and preferences.

The approach taken in this paper is agnostic in the sense that we simply correlate existing and important measurement systems as they are. We think this is an important exercise but it can only be a first step. What is needed is the development of a comprehensive framework that combines insights from the approaches taken by economists and psychologists to capture sources of heterogeneity in behavior. It is surprising that the Big Five apparently misses important preferences such as attitudes towards risk and time. Likewise the economic model is incomplete with respect to important preferences, but also with respect to capturing heterogeneity in abilities and beliefs. In the standard economic framework, beliefs are assumed to be endogenous to the strategic situation and formed in a rational way. Perhaps, with the exception of interpersonal trust, beliefs are typically assumed to follow common prior assumptions and rational updating. The importance of Locus of Control for explaining fundamental life outcomes on top of preferences, however, reveals the importance of enduring and individual specific belief systems. Other examples comprise optimism and pessimism, religious beliefs or ideological beliefs. The stability of belief heterogeneity is not well understood. It probably originates in different priors inherited from parents, self-selection into peer groups and institutions with reinforcing belief characteristics as well as boundedly rational belief formation, such as selected perception, non-Bayesian updating or ego utility (Koeszegi, 2006). Regardless of the precise channels that support enduring heterogeneous beliefs, economics would largely benefit from measuring and including them in explaining economic outcomes. In addition, economists have started to model the fact that preferences and beliefs are intimately related and not as separable as traditionally assumed. In fact, people often want to believe certain things, e.g., in terms of being liked by others or being better than others (overconfidence). Finally, another important extension of the economic model would be the measurement of person specific abilities. While IQ has become a standard individual specific characteristic to be included in outcome regressions, little work has acknowledged the importance of other competencies captured by Big Five traits, e.g. the role of conscientiousness for educational or labor market outcomes.

Chapter 2

Interpreting and decomposing the effect of non-cognitive skills on educational outcomes

2.1 Introduction

Multiple traits matter for success in life. Yet, the underlying dimension, classification, and identification of these traits are widely contested. For many economic models, a one dimensional skill or ability differentiates workers (Becker, 2009; Herrnstein and Murray, 2010; Neal and Johnson, 1996; Carneiro, 2003). A large literature studies how single ability, commonly not fully observed, affects educational choice and labor market outcomes (Heckman, 1979; Willis and Rosen, 1979; Card, 2001). More recently, economists have become interested in the multidimensional set of abilities which affect educational choices, educational success, and later life outcomes (Heckman and Rubinstein, 2001; Jacob, 2002; Heckman et al., 2006; Cunha and Heckman, 2007, 2008; Conti and Heckman, 2010). Much of this work adds a second "non-cognitive" or "socio-emotional" component which is an aggregate of skills or traits other than cognition that matter in life.

The measures used to identify the second component, also due to data availibility, vary widely. Some papers use revealed behavior at young ages, while others use responses to various questionnaires. Which questionnaires are used to estimate non-cognitive skills also varies across papers. For example, Heckman et al. (2006) use questionnaires on self-esteem and locus of control, while Jacob (2002) uses the portion of academic performance (grades, hours spent on homework) not captured by cognitive ability and disciplinary records from school. Conti and Heckman (2010) also use locus of control, but supplement this with measures of perseverance, cooperation, completeness, attentiveness and persistence. Heckman et al. (2013) allow the cognitive and the non-cognitive factor to load on 9th grade GPA and include measures of early risky behavior. Typically the cognitive and non-cognitive factors are extracted through factor analysis or through constructing indices. While the cognitive factor relatively clearly maps into the concept of intelligence[1], the extracted non-cognitive component must be interpreted by the factor's measurement system. The single "non-cognitive" trait is generally difficult to interpret and does not easily map into pre-existing taxonomies. Heckman et al. (2006) state in this context "we choose these measures because of their availability in the NLSY79. Ideally, it would be better to use a wider array of psychological measurements and ... to connect them with more conventional measures of preference parameters in economics." (p. 429). Using GSOEP data (Wagner et al., 2007), we find that the choice of which 2-factor system to estimate greatly influences what is actually measured and which conclusions are reached about the role of non-cognitive factor concerning important educational outcomes. Our results suggest a more careful interpretation within the debate on the importance of non-cognitive skills is needed.

Some papers that study cognitive and non-cognitive skills do not explicitly use a 2-factor structure (for example, Cobb-Clark and Tan, 2011; Farkas, 2003; Lleras, 2008; Rustichini et al., 2012). Rather, these papers include a number of measures which they believe proxy for non-cognitive skills and discuss how these measures affect the outcome of interest.[2] While these papers do not explicitly use a two fac-

[1]Psychological theory sometimes distinguishes between different components as fluid and crystalline intelligence, see e.g. Cattell (1987).

[2]This approach places less structure on the model, but also fails to account for measurement error (which can be corrected for when explicitly working with factors).

tor model, the papers vary widely in what they use for non-cognitive measures. For example, Kaestner and Callison (2011) use self-esteem and cognition to predict later life health, but do not find other non-cognitive measures predictive. Waddell (2006) similarly uses self-esteem and poor attitude early in school and finds that they affect later educational choices. Alternatively, Dunifon et al. (2001) use cleanliness and keeping an organized household as measures of non-cognitive ability using PSID data. Lleras (2008) explores how behaviors such as social skills, work habit, and participation in extracurricular activities directly predict later educational success and earnings. These papers vary widely in what traits they use to proxy non-cognitive skills and how these proxies are measured which makes it difficult to compare their findings.

Personality psychologists use a variety of personality taxonomies to predict and explain behavior. The most widely used model is the Big-5 personality inventory (conscientiousness, agreeableness, neuroticism, openness, and extraversion). The Big-5 (e.g. Costa and McCrae, 1992) were developed based on the lexical hypothesis by Allport and Odbert (1936), which suggests that individual personality differences are encoded in language (for an overview and discussion see Borghans et al. (2008)). In contrast, theoretical economic models include individual preference parameters which, while potentially related, are unique from ability. Developed as part of utility maximization theory, the two most common economic preference parameters are time preference[3] and risk preference (see e.g. Becker et al., 2012).

Measures of the Big-5 or economic preferences are absent from most economic surveys, limiting the possibility to integrate empirical and theoretical work and forcing empirical economists to develop ad-hoc non-cognitive factors. Despite their fundamental importance, little is known about how the non-cognitive factors estimated in the economics and education literature are related to the Big-5 and economic preference parameters. Moreover, non-cognitive factors are estimated using different measures and different methodologies across different studies, making it difficult to

[3]The economic concept of time preference is strongly related to concepts as self-control and self-regulation. For an overview see Frederick et al. (2002), for a discussion on the role of the different concepts in the development process see Bartling et al. (2010); Bettinger and Slonim (2007); Kosse and Pfeiffer (2012).

compare and interpret results.[4] Understanding how different estimated non-cognitive factors map into established taxonomies can help create consilience across previous work and can inform policy.

Previous literature has shown that school performance and education decisions are critical conjectures for later life outcomes (see e.g. Heckman et al., 2013). While there is a broad consensus that non-cognitive traits play a critical role in the determination of these outcomes and decisions (see discussion in Almlund et al., 2011), little is known about the actual drivers behind the abstract concept of non-cognitive traits. A deeper understanding about the actual processes might aid policy interventions in the context of educational decisions. There is a long lasting discussion on how to improve college access for children from families with a low socio-economic status. Van der Klaauw (2002) and Nielsen et al. (2010) explore the effect of financial aid while Bettinger et al. (2012) explore the effect of the assistance and information provisions. Knowledge about specific traits that drive the college enrollment decisions could lead to more targeted interventions. For example, if conscientiousness plays a crucial role, implementing interventions that are targeted to improve children's conscientiousness might be useful to improve college enrollment among certain subgroups.

In this paper, we use data from the German socio-economic panel (GSOEP) (Wagner et al., 2007), a unique panel data set which includes broad information about teenage behavior, preferences, personality, and later-life outcomes like GPA and college enrollment.[5] We construct several 2-factor models (cognitive and non-cognitive) based on measurement systems which are used in the previous literature. The resulting stylized models build the ground for a broad comparison of previously used non-cognitive factors. By decomposing these non-cognitive factors into a combination of underlying personality traits and economic preference parameters, we shed light on how non-cognitive models in the literature are related to each other and to

[4]Rustichini et al. (2012) and Becker et al. (2012) relate economic preferences to personality measures, but do not try to decompose non-cognitive skills.

[5]Germany is an ideal country to study due to very low college fees and broadly available student grants (BAföG). Students face relatively few financial constraints which could otherwise bias the decision to go to college.

more traditional taxonomies. In doing so, we create a road map of which personality traits and preferences are measured when using different identification strategies for the non-cognitive factor. We find that the choice of measures used to construct the non-cognitive factor can broadly change its interpretation. The second half of this paper compares the 2-factor models to our "preferred" model, which uses the Big-5 and economic preferences directly. We find that the preferred model outperforms all of the 2-factor models in predicting educational success. Moreover, there is a great deal of heterogeneity between non-cognitive factors. The more correlated the non-cognitive factor is with traits from our preferred model, such as conscientiousness and time preference, the more predictive it is of later outcomes. The combination of the models shows that the non-cognitive factors add little to no predictive power to the preferred model.

2.2 Data and measures

Our analysis uses data from the youth questionnaire of the GSOEP which is conducted at age 17, the age at which children first answer questionnaires on their own. Starting in the year 2006, the youth questionnaire measures preference and personality and includes a short IQ test. We complement this information with revealed behaviors and outcomes collected in following waves. Our main analysis uses data on more than 1,300 adolescents interviewed at age 17 in the years 2006 to 2012.

The Big-5 inventory is measured using a validated 15-item questionnaire (Gerlitz and Schupp, 2005) that is commonly used in empirical personality research (see e.g. Becker et al. (2012)). Risk preference is measured by the question "How do you see yourself: Are you generally a person who is fully prepared to take risks, or do you try to avoid taking risks?" Answers were given on an 11-point scale, where zero means "unwilling to take risks" and 10 means "fully prepared to take risks" This so called general risk question has been studied in various papers and is highly correlated with incentivized experimental measures and revealed behavior (see e.g. Dohmen et al. (2011)). To measure time preference, the participants rate how strongly they agree with the two statements "I abstain from things today to be able to afford more tomorrow" and "I prefer to have fun today and don't think about tomorrow" (reversed) on a 7-point Likert scale. We construct our measure of time preference by summing the standardized responses. The resulting score is well correlated with incentivized experimental measures of time preference.[6] To measure cognitive skills, the participants took part in a validated short version of the well-established "I-S-T 2000 R" (Amthauer et al., 1999), covering all three subsets which are verbal, numerical and figural abilities (for details see Solga et al. (2005)).

We estimate four different stylized 2-factor models where the non-cognitive factor in each model is constructed using different reported behaviors or questionnaire responses. The different measurement systems used to identify the non-cognitive factor are chosen to be similar to measurement systems used in the economics and education literature. The cognitive factor is uniformly constructed from three IQ sub-tests across all models.[7] Each model estimates the two factors jointly by confirmatory factor analysis and allows for correlation between factors. We assume that the factors have multivariate normal distributions, though our conclusions do not change when using minimum-distance estimators which do not rely on the assumption of normality. For each 2-factor system, a different set of behaviors or responses are used to identify the non-cognitive factors: Model 1 (NC-LOCUS) uses responses to a 10-item Locus of Control questionnaire (Rotter, 1966), as has been done in work using the National Longitudinal Survey of Youth 1979, such as in Heckman et al. (2006). Model 2 (NC-ENGAGEMENT) uses participation in extra-curricular activities, time used on productive or enriching tasks, time used on unproductive or passive tasks, and number of close friends. Model 3 (NC-RELATIONS) uses responses on if the individual argues with their parents, argues with their friends or significant others, the number of close friends which they have, and responses on communication and interactions between the individual and their parents. Model 4 (NC-BEHAVIORS) uses reported behavior on drinking habits, smoking habits, eating habits, if the individual argues with their parents, and if the individual ar-

[6] E.g. using data of Vischer et al. (2013) indicates a highly significant correlation ($p < 0.001, N = 965$).

[7] Some papers use achievement tests to measure cognition rather than IQ. As shown in Borghans et al. (2011), achievement tests may capture traits other than IQ. As the GSOEP does not include achievement test scores, we cannot consider variation along this dimension.

gues with their friends or significant others.[8] Across the four models, we believe we have embodied many of the identification strategies used by work on the importance of non-cognitive skills. We compare each 2-factor model to our "preferred" model, which uses the Big-5 and economic preference factors directly, either with IQ (Pref-1) or without IQ (Pref-2). For an overview of the four 2-factor models and our preferred models, see Table 2.1.

Table 2.1: Measurement systems of different non-cognitive constructs

Model	**Measurement System**
NC-LOCUS (NC-L)	Rotter's Locus of Control (10-items).
NC-ENGAGEMENT (NC-E)	Frequency of engagegment (volunteering, sport, technical work, reading), number of close friends.
NC-RELATIONS (NC-R)	Relation to parents and friends (bonding, love, argues or fights, problems solving), number of close friends.
NC-BEHAVIORS (NC-B)	Consumption behavior of alcohol and tabacco, eating behavior, argues or fights with family or friends.
PREFERRED incl. IQ (Pref-1)	Big-5 (conscientiousness, agreeableness, neuroticism, openness, extraversion), economic preferences (risk and time), IQ.
PREFERRED w/t IQ (Pref-2)	Big-5 (conscientiousness, agreeableness, neuroticism, openness, extraversion), economic preferences (risk and time).

2.3 Results

We find that different constructions of the non-cognitive factor lead to different conclusions about the relative importance of non-cognitive skills. The extracted non-cognitive factors have only moderate (or even negative) correlations across models and have plausible but varying correlations with the Big-5 and economic preference

[8] There are only around 750 observations for this measure as it uses variables which are collected in later questionnaires, which have not yet been answered by all those who have answered the age 17 questionnaire.

factors. We decompose each non-cognitive factor by regressing them on the covariates from our preferred model. We find that different personality traits and economic preferences play important roles in the different constructions of the non-cognitive factors. Finally, we compare regressions of GPA and college enrollment on each of our 2-factor models and on our preferred model. We find that the non-cognitive factors vary widely in their ability to explain outcomes, that the preferred model outperforms every 2-factor model, and that the non-cognitive factor from each 2-factor model adds little or no additional explanatory power to the preferred model.

As shown in Table 2.2, the correlations between the four non-cognitive factors are quite low. Moreover, different non-cognitive factors are correlated with different aspects of the Big-5, economic preference parameters, and IQ. The correlation between the different non-cognitive factors ranges between -0.12 and 0.20. Interestingly the correlation between NC-LOCUS and NC-BEHAVIORS is negative, while the correlations between NC-RELATIONS and NC-LOCUS and between NC-RELATIONS and NC-BEHAVIORS are positive. This suggests that each factor may be capturing different aspects of a vector of unobservable non-cognitive characteristics. NC-LOCUS is correlated with conscientiousness and IQ and has a strong negative correlation with neuroticism. NC-ENGAGEMENT has lower correlations, but is correlated most with risk, openness, and extraversion. NC-RELATIONS is highly correlated with many of the other factors. It is positively correlated with conscientiousness and openness. NC-BEHAVIORS is positively correlated with both conscientiousness and neuroticism but negatively correlated with risk preference, extraversion, and IQ.

Table A2.1 regresses the cognitive factor and each of the non-cognitive factors on the Big-5 personality traits and economic preferences.[9] This table provides similar evidence as Table 2.2, but now considers partial correlations between parameters in our preferred model and the non-cognitive factors. For each factor we provide the regression on only the Big-5 and on both the Big-5 and economic preference parameters. First, we find that cognition is correlated with personality traits. As found in the psychology literature, we find the strongest partial correlation with open-

[9]In actuality, each model has a uniquely estimated cognitive factor as the factors in each model are estimated jointly. Yet, the cognitive factors are estimated using the same measures across models and have correlations of nearly unity.

Table 2.2: Correlations between different noncog. and cog. constructs

	NC-L	NC-E	NC-R	NC-B
NC-L	1			
NC-E	0.111***	1		
NC-R	0.198***	0.0968***	1	
NC-B	-0.122***	-0.0367	0.0844**	1
Cons.	0.255***	0.0985***	0.192***	0.130***
Agree.	-0.148***	-0.0253	-0.224***	-0.0660*
Neuro.	-0.459***	-0.0243	-0.0968***	0.123***
Open.	0.0848***	0.185***	0.180***	-0.0552
Extra.	0.180***	0.127***	0.154***	-0.149***
Time	0.0977***	0.0741***	0.123***	0.0974***
Risk	0.0761***	0.0952***	-0.0320	-0.186***
IQ	0.296***	0.122***	0.144***	-0.136***

Notes: Table shows correlations between the constructed non-cognitive factors, the Big 5 personality traits, discount rate, risk aversion, and IQ. NC-L is based on the Rotter's Locus of control. NC-E is based on engagement behavior, NC-R is based on self-reported relationships, and NC-B is based on self reported risky behaviors. Number of observations varies between 760 and 1416 due to data availability. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ness. Agreeableness is also positively correlated with the cognitive factor, while extraversion and neuroticism are negatively correlated. Considering the non-cognitive factors, we see that the ceteris paribus relationship with neuroticism tends to be negative and statistically significant, but that the relationships vary for other parameters. Conscientiousness tends to be positively related, but is not significant in some models. Agreeableness is positively associated with some non-cognitive factors but negatively associated or unassociated with others. Similarly, risk preference varies between positively related, unrelated, and negatively related with the non-cognitive factor. Evaluating the non-cognitive factors according to the R^2 by personality and preference measures reveals fundamental differences. While Big-5 and preference measures can explain more than 25% of the variation in NC-LOCUS, they explain only about 5% of the variation in NC-Engagement.

Given the importance of non-cognitive skills in the determination of school performance and education decisions (see discussion in Almlund et al. (2011)), we will focus on educational success and decision making as our key outcomes in the remaining analysis. Tables 2.3 and A2.2 consider how different 2-factor models predict GPA and college enrollment decisions. Both models control for gender, urban status, and residence in Eastern Germany. The regression model in Table 2.3 also controls for the secondary education tier in which the grade was received.[10]

The dependent variable in Table 2.3, GPA, can range from one to six and is coded such that higher values indicate better performance. Due to data availability, GPA is regarded at age 17 and calculated as the average of grades in mathematics, German and first foreign language. The mean (standard deviation) in our sample is 4.07 (0.72). First, we regress GPA on the respective 2-factor models, then these results are contrasted to those of our preferred model (full set of Big-5, economic preferences and IQ), and finally we check if the extracted factors provide additional predictive validity over our preferred model. We find that cognitive ability predicts GPA, but that so do all of the non-cognitive traits (though they vary in size and significance). The smallest statistically significant coefficient is less than half of the size of the largest coefficient. Turning to the preferred model, we see that cognition, conscientiousness, agreeableness, and time preference are positively correlated with GPA while risk preference is negatively correlated. The preferred model explains six percent[11] more of the variance compared to the 2-factor models.[12] When we include the different non-cognitive factors in the preferred model, none provide substantial additional explanatory power over the preferred model alone.

[10]The secondary education system within Germany has basically three tracks (low, medium, high) which are supplemented by comprehensive schools and vocational school.

[11]Although, the compared models vary regarding the number of coefficients, due its straightforward interpretation our analysis is based on comparisons of R^2 instead of e.g. adjusted R^2 or information criteria. Given the large number of observations the results remain very similar when using other measures of fit.

[12]Note that personality and economic preference have substantial predictive power and explain 12% of the variance in GPA without the inclusion of IQ.

Table 2.3: Model comparison: GPA

	NC-L	NC-E	NC-R	NC-B	Pref-1	Pref-2	Comb-L	Comb-E	Comb-R	Comb-B
Cog	0.246***	0.253***	0.249***	0.273***	0.234***		0.238***	0.233***	0.231***	0.246***
	(0.019)	(0.019)	(0.019)	(0.025)	(0.020)		(0.020)	(0.020)	(0.020)	(0.027)
Noncog	0.038**	0.031*	0.055***	0.070***			-0.012	0.021	0.029	0.042*
	(0.019)	(0.018)	(0.018)	(0.025)			(0.021)	(0.018)	(0.019)	(0.025)
Cons.					0.170***	0.152***	0.171***	0.168***	0.168***	0.177***
					(0.022)	(0.023)	(0.022)	(0.022)	(0.022)	(0.030)
Agree.					0.054**	0.087***	0.053**	0.052**	0.059***	0.072**
					(0.021)	(0.022)	(0.021)	(0.021)	(0.021)	(0.028)
Neuro.					-0.038*	-0.109***	-0.043*	-0.038*	-0.037*	-0.031
					(0.021)	(0.022)	(0.023)	(0.021)	(0.021)	(0.029)
Open.					0.025	0.184***	0.023	0.021	0.025	0.011
					(0.034)	(0.034)	(0.035)	(0.035)	(0.034)	(0.047)
Extra.					-0.024	-0.147***	-0.022	-0.023	-0.027	0.018
					(0.033)	(0.033)	(0.034)	(0.033)	(0.033)	(0.046)
Risk					-0.062***	-0.074***	-0.061***	-0.064***	-0.060***	-0.067**
					(0.019)	(0.020)	(0.019)	(0.019)	(0.019)	(0.027)
Time					0.024	0.040**	0.024	0.023	0.023	0.036
					(0.019)	(0.020)	(0.019)	(0.019)	(0.019)	(0.026)
Observations	1327	1327	1327	732	1327	1327	1327	1327	1327	732
R^2	0.145	0.144	0.148	0.147	0.208	0.121	0.208	0.209	0.209	0.220

Notes: Table shows regressions of GPA on one of the four constructed 2-factor models, our two preferred models, or combined models. NC-L is based on the Rotter's Locus of control. NC-E is based on engagement behavior, NC-R is based on self-reported relationships, and NC-B is based on self reported risky behaviors. All estimated OLS models include the following controls: gender, urban status, residence in Eastern Germany and the education tier in which the grade was received. Standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2.2 mirrors Table 2.3, but considers college enrollment. We consider college enrollment at age 21 and only include those who reached age 21 by the year 2012 and were in the top tier of the secondary education system.[13] Of those, 51.1% enrolled into college.[14] Table A2.2 displays the average marginal effects on the choice to enter college.[15] Unlike GPA, we find that while the cognitive factor matters, none of the non-cognitive factors play a statistically significant role.[16] Yet, when we consider the preferred model, we find that it outperforms the 2-factor models in terms of pseudo R^2 by a factor of 1.5, with conscientiousness playing a statistically significant role in the choice to enroll in college. In contrast to the small and statistically insignificant non-cognitive factors, the coefficient on conscientiousness is nearly as large as the coefficient on cognition and is statistically significant. This indicates a substantial contrast between the ad-hoc two factor models and the preferred model.[17]

2.4 Conclusion

This paper aims to compare many different identification strategies for a non-cognitive factor and to decompose and interpret their relative effectiveness. We construct stylized factors based on identification strategies in the previous literature and relate the different estimates of non-cognitive skills to each other and to established taxonomies. The correlation between the different non-cognitive factors varies greatly, ranging between -0.12 and 0.20. Given the heterogeneity among the different of non-cognitive factors, we compare them with a preferred model which contains

[13]Only graduation in the top tier of the secondary education gives you full access to the college systems. For those individuals who spent a year of civilian or military service we consider college enrollment at the age of 22.

[14]We consider all types of college, including universities and universities of applied sciences.

[15]Note this is the average of the marginal effects rather than the marginal effect at the average.

[16]Lack of statistical significance may be driven by the fewer observations – due to restrictions on school leaving degree and age cohort.

[17]To rule out that the relations between personality factors and educational outcomes are purely driven by heterogeneities in the socio-economic background we repeat the analysis of Table 2.3 an Table A2.2 and include parental education as an additional control. The results are displayed in Appendix A2. As expected, parental education has a significant positive relation to child's GPA and college enrollment decision, but it only explains 2% more of the variation in the data. The effect sizes of the cognitive, non-cognitive, and personality factors remain the same magnitude, though they are generally smaller than when parental education is omitted.

personality traits and economic preference parameters. The Big-5 and economic preferences explain between 5% and 25% of the variation in different non-cognitive factors.

In the context of educational outcomes, we find that the preferred model outperforms all of the 2-factor models in terms of predictive power. Moreover, there is an enormous heterogeneity among the relevance of the different non-cognitive factors. In general, the higher the correlation with traits from our preferred model, such as conscientiousness and time preference, the higher their predictive power concerning educational outcomes. The combination of the models indictes that the ad hoc measured non-cognitive factors add little to no predictive power to the preferred model, which indicates that the underlying non-cognitive traits of the ad hoc measured models are already part of the measurement system of the preferred model.

In summary, the choice of which 2-factor system to estimate greatly influences what is actually measured and what conclusions are reached about the role of non-cognitive skills in life – and its relative importance compared to cognitive skills. Our results suggest a more careful interpretation within the debate on the importance of non-cognitive skills is needed.

Chapter 3

Breastfeeding duration, early life circumstances and the formation of human preferences

3.1 Introduction

Human preferences such as time, risk and social preferences are key building blocks of any economic model and fundamentally determine human behavior and life-outcomes (Heckman et al., 2006; Sutter et al., 2013; Dohmen et al., 2011; Becker et al., 2012). For example, time preferences are relevant for any type of investment decision because investment, by its nature, requires patience, i.e., a willingness to forego current consumption and wait, in order to earn a higher return in the future. Similarly, most decisions involve uncertainty, and risk preferences determine how someone behaves in the presence of uncertainty. Moreover, almost all social interactions are shaped by social preferences, as e.g., altruism.

Despite their fundamental importance, little is known about how human preferences form. Understanding the process of preference formation is of great relevance not only for behavioral sciences that model behavior based on preferences, but also for informing policy as it helps to uncover reasons for social mobility and provides insights concerning the effectiveness of early childhood interventions. To shed light on the process of preference formation, we focus on early childhood which has been indicated as critical and sensitive period in the human development process by various fields of research (Heckman, 2006; Cunha and Heckman, 2007; Currie and Almond, 2011; Lanigan et al., 2009; Ainsworth and Bowlby, 1991; Sroufe, 1988). First, we present evidence for breastfeeding duration as valid measure of quality of early life circumstances and quality of parenting in general. Building on these findings, in our main analysis, we secondly investigate the role of early life circumstances in systematically shaping human preferences.

3.2 Breastfeeding and quality of early life circumstances

The role of breastfeeding and associated effects have been studied in various fields (Heinrichs et al., 2002; Meyer-Lindenberg et al., 2011; Uauy and De Andraca, 1995; Chen and Rogan, 2004). In this section, we provide new evidence for the convergent and discriminant validity of breastfeeding duration as a measure of quality of early life circumstances. Note that the extent to which breastfeeding duration is a measure of quality of early life circumstances varies between cultures. Our validation concerns Western societies, especially Germany. In developing countries the decision to breastfeed may be determined by very different considerations (Jayachandran and Kuziemko, 2011). Our analysis is based on data from the SOEP (SOEP, 2012; Wagner et al., 2007), a large representative panel of the German population including very detailed information about life circumstances in the first three years of life for about 550 children of the birth cohorts 2004-2007 and their natural parents. The share of children who were ever breastfed is 88%, of these, the average duration of non-exclusive breastfeeding is 7.6 months (std. dev. is 6.0 months). Figure A3.1 displays the cumulative distribution of breastfeeding duration and reveals substantial variation. In analyzing this data we find that the decision to breastfeed at all is mainly driven by maternal health and socio economic status. In contrast, the variation in breastfeeding duration reflects heterogeneities in the quality of time spent with children, i.e., of early life circumstances.

When their children are between two and three year years old, mothers are asked how many times in the last 14 days she, or the main caregiver, has done nine particular activities together with their child. A principal component analysis concerning responses to all potential activities yields three components (see Table A3.1). The first component reflects activities which involve face-to-face contact and intense interaction between mother and child such as reading or telling children's stories or singing children's songs with the child (high quality time). The second component reflects activities with less intense interaction and direct contact such as going shopping or visiting other families with the child (medium quality time). The third component comprises watching TV or videos (low quality time).

In Table A3.2 we report results of multivariate regression analyses to investigate determinants of being breastfed at all (Column 1) and of breastfeeding duration given a child was breastfed (Column 2). Results in Column (1) indicate that being breastfed at all is strongly determined by physical health problems of the mother ($p < 0.01$) as well as socio-economic status ($p < 0.05$) while there is no jointly significant relation to the quality of time spending (see bottom of Table A3.2 for Wald-tests). In contrast, results of the multivariate analysis in Column (2) reveal that, breastfeeding duration is highly positively correlated with high quality time spending ($p < 0.01$). The correlation is weakly negative for spending medium ($p < 0.1$) and low quality time ($p < 0.1$). The duration is not significantly correlated with health conditions of the mother or maternal education and household income. Using other measures of positive early childhood environment and other data sets confirms these results. Table A3.3 reports positive correlations of breastfeeding duration with quality of stimulations and support using the HOME inventory (Bradley and Caldwell, 1981; Blomeyer et al., 2009), life satisfaction of the mother in the year of the child's birth and how important it is for the mother to have children.

In sum, the analysis justifies the interpretation that the breastfeeding duration is determined by an underlying and not directly observable quality of early life circumstances and an error term, which could reflect, e.g., local breastfeeding traditions or health shocks. If this error term is uncorrelated with the underlying quality of early life circumstances (classical error-in-variables assumptions) (Wooldridge, 2010) the estimation coefficients on breastfeeding duration, represent a lower bound estimate of the effect of quality of early life circumstances. To limit the attenuation bias and since never being breastfed is to a large extent determined by health shocks, we will restrict the analysis to children who were breastfed.

Using breastfeeding duration as a measure of early life circumstances offers a great potential: Research has shown that mothers remember breastfeeding durations in a valid and reliable way (Li et al., 2005). Therefore, information about breastfeeding duration is accessible in retrospect, enabling researchers to explore the effect of early life circumstances in various independent samples covering different age cohorts and outcomes.

3.3 Early life circumstances and preference formation

The validation of breastfeeding duration as a measure of quality of early life circumstances sets the stage for our main analysis. We study the effect of early life circumstances on the formation of fundamental preferences in two independent data sets. The first is a preschool children sample that allows analyzing the effect of interest as close as possible to children's early childhood experience. The complexity of experiments to elicit preferences makes studying even younger children difficult. The second data set repeats the setting for a sample of young adults. Comparing our findings between these two data sets enables us to study robustness, and whether the observed pattern persists into adulthood.

Data set 1: The preschool sample consists of 302 breastfed children and their natural mothers (see Methods for details on design of the study and experimental procedures). 194 mother-child pairs took part in a time preference experiment while 108 pairs took part in experiments to measure risk preferences and altruism. All experiments were run using real incentives and all children had to answer control questions to check understanding of experimental instructions and payment rules. Children who were unable to properly answer the control questions were excluded from the analysis. To measure children's time preference we used an adaption of the Marshmallow Experiment (Mischel et al., 1989; Bartling et al., 2010). In this task children face the trade-off between receiving a smaller reward (one pack of gummy bears) sooner, or waiting and receiving a larger reward (two packs of gummy bears) later. In the risk and social preference experiments children could earn chips, which were exchanged for toys at the end of the experiment. To measure children's risk preferences, they played a version of the so-called devil's task (Slovic, 1966). To elicit altruism, a simple distribution choice was used (Fehr et al., 2008). In particular, the child had to decide between allocating either two chips to himself and no chip to another anonymously matched child or to allocate one chip to himself and one chip to the other child. In addition, IQ of the children was measured accounting for crystalized and fluid aspects of intelligence (Cattell, 1971).

To measure mothers' time and risk preferences as well as altruism, standard incentivized experimental protocols were used (Dohmen et al., 2010). In addition, mothers completed a short IQ Test and answered a detailed survey, covering items such as personality of the mother (Big-5), socio-economic background and breastfeeding duration.

On average, the children are 72.6 months (5.95 years) old, 51.0% are male and mean age of mothers is 37.0 years. Our main result is displayed in Panel A of Table 3.1. The dependent variable is the child's respective preference, which is regressed on breastfeeding duration. Columns 1, 3 and 5 show the estimates without further controls. The results indicate that a longer duration of breastfeeding is associated with a lower willingness to take risk and higher levels of patience and altruism. Columns 2, 4 and 6 include controls which might affect breastfeeding duration and preference formation simultaneously (Jayachandran and Kuziemko, 2011; Currie and Moretti, 2003). These controls include child characteristics (e.g., gender and IQ), socio-economic family background (e.g., income and education) as well as personality and other characteristics of the mother. For example, if more patient and educated mothers breastfed longer, the omission of a mother's time preference and education would potentially lead to an overestimation of the effect of breastfeeding duration on a child's patience. Furthermore, we include controls specific to the respective experiment, such as elapsed time since the last bigger meal in the time preference regression. Results in columns 2, 4 and 6 show that the observed pattern is very robust to including these controls. A comparison of the coefficients of breastfeeding duration to those of the control variables (see Table A3.4) further reveals the quantitative relevance of early life circumstances. To illustrate, a simulated change in breastfeeding duration from the 25%-percentile (2 months) to the 75%-percentile (9 months), increases the probability of being patient by 18.5% points (from 68.6% to 87.1%), which, e.g., exceeds the average effect of a two standard deviations increase in mother's patience.

Wald tests reveal significant joint effects of child characteristics as well as personality and preferences of the mother on the child's preferences, but no significant effect of the socio-economic environment (see Table 3.1). In sum these results suggest that the quality of early life circumstances and personality of the mother play a more crucial role in the process of preference formation than intellectual or monetary resources.

Data set 2: To study robustness of the reported pattern and persistence into adulthood we study a sample of 175 breastfed university students. Mean age of students is 21.8 years and 44.6% of them are male. Experimental preference measures were obtained using standard and well-established tools for risk and time preferences (Becker et al., 2012; Dohmen et al., 2010). Altruism was elicited using a 3-item questionnaire. We repeat the same analysis as for the preschool sample and report the results in Panel B of Table 3.1. Columns 1, 3 and 5 display estimations without controls. In Columns 2, 4 and 6 we include controls for important individual characteristics such as gender, age, and intelligence, as well as occupation of the parents. The signs of the coefficients of interest are identical to the preschool sample. Longer breastfeeding duration is statistically significantly associated with less willingness to take risks and higher levels of patience and altruism. As before, all effects are robust to including a set of controls. In terms of effect size the coefficients concerning the two samples are difficult to compare due to the different sample compositions and

	Time		Risk		Altruism	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A preschool children						
Dependent variable	Binary		Standardized		Binary	
Type of estimation	Probit		OLS		Probit	
Duration of breastfeeding (in months)	0.025***	0.024***	-0.032**	-0.033*	0.016**	0.021***
	(0.008)	(0.008)	(0.016)	(0.020)	(0.007)	(0.006)
Individual characteristics	No	Yes	No	Yes++	No	Yes++
Socio-economic environment	No	Yes	No	Yes	No	Yes
Personality/preferences/IQ of mother	No	Yes	No	Yes	No	Yes++
Task specific controls	No	Yes	No	No	No	Yes
Observations	194	194	108	108	100	100
(Pseudo) R-squared	0.047	0.179	0.025	0.183	0.058	0.286
Panel B young adults						
Dependent variable	Standardized		Standardized		Standardized	
Type of estimation	OLS		OLS		OLS	
Duration of breastfeeding (in months)	0.042**	0.038**	-0.039*	-0.047**	0.047***	0.044***
	(0.016)	(0.016)	(0.023)	(0.020)	(0.015)	(0.015)
Individual characteristics	No	Yes+	No	Yes	No	Yes
Socio-economic environment	No	Yes+++	No	Yes+++	No	Yes++
Observations	175	175	175	175	175	175
R-squared	0.028	0.222	0.021	0.210	0.037	0.165

Table 3.1: The effect of quality of early life circumstances on preferences of preschool children and young adults. Displayed coefficients are average marginal effects with respective preference as dependent variable and robust standard errors in parentheses. Panel A shows results for the sample of preschool children, Panel B for the sample of young adults, respectively. The complete specifications and estimation results can be found in Tables S4 and S5. ***, **,* indicate significance at 1-, 5-, and 10-percent level, respectively. +++, ++, + indicate significance at 1-, 5-, and 10-percent level, of Wald-tests testing the hypothesis that all coefficients of the respective category are zero.

elicitation techniques. However, concerning risk preferences where we use ratio scale measures in both samples, effect sizes are moderately bigger for the young adults than for the preschoolers. This suggests an increasing impact of early life circumstances within the development process which is in line with a self-productive and dynamic complementary pattern of a production function (Heckman, 2006; Cunha and Heckman, 2007).

3.4 Real-life behavior and cohort effects

We conclude with answering two important questions arising from our main finding: (i) how does the effect of early life circumstances on preferences translates into real-life behavior or outcomes, and (ii) is the long-term variation in breastfeeding practices that is characteristic for western societies related to heterogeneities in preferences across birth cohorts? Concerning the first question, note that previous work has shown positive behavioral and health outcomes for more patient and risk averse individuals (Sutter et al., 2013; Ida and Goto, 2009). These individuals forego immediate risky pleasures such as smoking, drinking or eating sweet food, which decreases the risk of adverse health effects later in life. In light of these findings we hypothesize that because a higher quality of early life circumstances is associated with higher levels of patience and less willingness to take risk, a higher quality of early life circumstances should also positively affect health and self-control outcomes. Using the German Health Survey for Children and Adolescents (KiGGS, 2008) we find supporting evidence for this hypothesis. In particular, we analyze health-related behaviors and outcomes, i.e., body-mass index (BMI), smoking and drinking behavior of more than 4,000 breastfed adolescents at the age of 11 to 17 years and relate these behaviors to breastfeeding duration. As hypothesized, results in Table A3.6 reveal that breastfeeding duration is negatively associated with BMI and the probability of smoking and drinking. These relations are robust to controlling for socio-economic status of the families.

To address the second question, note that breastfeeding durations have undergone substantial historical variations implying that there are not only heterogeneities in breastfeeding duration within a birth cohort but also across birth cohorts (Heimerdinger, 2009). Although there exist no administrative data on breastfeeding in Germany for birth cohorts before 1986, it is possible to construct historical breastfeeding patterns. In particular, we use breastfeeding data recorded for the purpose of marketing research by Nestlé. The data consist of breastfeeding quotas and breastfeeding durations elicited in interviews of 250 mothers (each wave) during the 1970s to the 1990s (see Figure A3.2). We match this data by year of birth with preference data from the SOEP (SOEP, 2012; Wagner et al., 2007) and construct a panel on cohort level including repeated measures of preferences (see Methods for details). Estimates for each preference are reported in Table A3.7. It turns out that the coefficients of interest mirror the patterns we found in our two cross-section data sets on an individual level. In a historical perspective and on a cohort level, breastfeeding duration is again negatively associated with willingness to take risk ($p < 0.05$) and positively associated with altruism ($p < 0.01$) and patience (not significant).

3.5 Concluding remarks

We have explored the role of quality of early life circumstances, measured in terms of breastfeeding duration, concerning the formation of preferences. Using two independent data sets, varying by age of participants (preschool children vs. young adults), we find a robust, systematic and persistent effect of early life circumstances on time, risk, and social preferences. Our results imply that early life circumstances, above and beyond purely intellectual or monetary resources, are a crucial determinant for the development of human personality with far-reaching implications for social mobility and early childhood intervention policies. We also report relevance of early life circumstances concerning health-related behavior and outcomes in a way that is predicted by the specific preference formation pattern observed in our two data sets. Using historical variations in the population as a whole we further document long-term effects on preference formation, empirically complementing theoretical contributions in this domain (Doepke and Zilibotti, 2008, 2012). Our finding concerning historical variations suggests an interesting interaction, which has not received much attention yet: changes in "technology", such as parental care, have the potential to systematically affect changes in preferences on a societal level.

3.6 Methods

Panel A: Preschool children

Sample. Data concerning preschool children consist of two independent sub-data sets which are administered by the DIW Berlin under the names MuKi III b and c. These data sets feature experimental measures concerning time preference (Part A), risk preferences (Part B) and altruism (Part B) run with preschool children and their mothers. All experiments and interviews were conducted by specially trained and experienced interviewers from the same organization that collects the data for the SOEP. The data sets include intelligence and personality measures as well as information on socio-economic background. The time preference experiment was conducted at the families' own homes (see below) while the experiments of the children in Part B were conducted in an extra room in children's day-care centers. The interviews of the mothers were generally held at their own homes. All mothers took part in a two-part computer assisted personal interview (CAPI) conducted with a laptop (Dohmen et al., 2010). In the first part mothers answered a detailed survey including demographic and socio-economic questions, as well as, questions concerning breastfeeding duration and their personality (all questions are based on the SOEP questionnaires) (Wagner et al., 2007). In the second part mothers took part in a short intelligence test and incentivized behavioral experiments.

All experiments were run using real incentives. In Part A children decided about gummy bears, while in Part B they could earn chips, which were exchanged for toys at the end of the experiment. In our analysis we only include children who demonstrated understanding of the experiment (control questions and interviewer rating). We exclude children from Part A for whom the mother indicated that her child does not like gummy bears at all or not so much. For non-biological children it is very unlikely that breastfeeding duration is a valid measure of quality of early life circumstances and therefore we also restrict the analysis to biological children.

To receive comparable results we also exclude observations with missing values in covariates from the analysis (see Table A3.4).

Experimental measures. The experiment concerning children's *time preference* was a field adaption of the "marshmallow experiment" of Walter Mischel (Mischel et al., 1989; Bartling et al., 2010) and was conducted at the families' own homes. Mother, child and the interviewer remained in the same room. Before the detailed interview of the mother started, the interviewer opened a pack of gummy bears and explained that the child could either eat them now or wait until the end of the mother's interview and receive an additional pack. Thus, children were faced with the decision between receiving a smaller reward (one pack of gummy bears) sooner, or waiting and receiving a larger reward (two packs of gummy bears) later. Gummy bears were used since they are more popular in Germany than marshmallows. 23.7% of the children took the opened pack before the interview ended. They are classified as impatient. 76.3% waited and received two packs of gummy bears. They are classified as patient.

Concerning their *risk preferences*, children played an adaption of the devil's task (Slovic, 1966). They were presented with 10 indistinguishable closed boxes of which nine included a chip and one a robber (in the English original it is a devil, the direct translation to the word used here is robber). The children could sequentially open as many boxes as they wanted to. They could keep all chips, which they found in the opened boxes, but if they opened the box with the robber, they lost all chips of this round. The game was played for six rounds and we use the average voluntary stopping point as measure for children's willingness to take risk. The average voluntary stopping point is not identified for those children who never stopped voluntarily and never had the chance to open the 9^{th} box because the robber always occurred before. For those three children we assume the voluntary stopping point to be the 9^{th} box (the maximum, upper bound) which seems plausible since they opened the 7^{th} or 8^{th} box when they had the chance to do so. The analysis shows very similar results if we use the maximum number of opened boxes (lower bound) for these children instead. The mean average voluntary stopping point is 5.08 (standard deviation 2.19).

To elicit *altruism*, we used a distribution choice as in the study of (Fehr et al., 2008). The child had to decide between different distributions of chips affecting himself and another anonymously matched child that could be from the own kindergarten group or from another unknown kindergarten. The child was told whether the receiver was from the own group or not and we control for the different setting in our analysis. We focus on the costly altruism variant of the game. In this game the children had to decide between two chips for themselves (2,0) or one for themselves and one for the other child (1,1). The 15.0% children who chose (1,1) are classified as altruistic. Similar as in Fehr et al., the children also played three other variants of the game.

Cognitive ability. The children took part in three modules of intelligence tests. Two of them are sub modules of the Culture Fair Intelligence Test Scale 1 (CFT1) (Weiss, 2006) and measure the fluid intelligence of children. The first submodule was a classification test where the child had to find one out of five symbols which does not fit into the row. The second one was a matrix test where the child had to add a fitting pattern to a row of three patterns. Both subtests contain 12 items. The sum of the correct answers builds the fluid intelligence score. The third module was a modified version of the German Peabody Picture Vocabulary Test Revised (PPVT-R) (Bartling et al., 2010), in which the child heard a word and had to match it to one out of four symbols. This test is a verbal scale, which captures culture and education related components of intelligence. The test contains 61 items and the number of correct answers reflects the score of crystallized intelligence of the child. To obtain an IQ score of the child, first both scores were standardized, then the standardized values were added up and finally the resulting sum was standardized again. This results in a score with mean equal zero and standard deviation equal one.

Our study also uses an intelligence test concerning the mother. Frieder Lang developed ultra-short tests concerning cognitive skills of adults; one of them is the Symbol-Digit-Test (SDT) (Lang et al., 2005), which is a modified submodule of the Wechsler Adult Intelligence Scale (HAWIE-R) (Tewes, 1994). In the SDT mothers had to match as many numbers and symbols as possible according to a correspondence list within 90 seconds. The results of this ultra-short test correlate well with test scores from well-established intelligence tests (Lang et al., 2005, 2007).

Experimental measures of preferences of mothers. All experiments concerning mothers preferences were conducted in their own homes as a part of a computer assisted personal interview (CAPI). Concerning time and risk preferences of mothers, the same procedures and protocols as in Dohmen et al. (2010) were used. Mothers were informed that in case of a win they would receive the amount as a cheque by mail. For measuring time preference mothers faced the trade-off between receiving 100 € "today" and receiving a higher amount in six months. The offered higher amount started at 101.2 € and was increased in 19 further steps of 2.5% p.a.. One out of seven mothers was randomly selected and was paid according to one of her decisions, which was selected randomly. This ensures incentive compatibility. We use the standardized reversed first switching row as measure for time preference of the mothers.

Mothers' *risk preferences* were measured in a similar way. Here they had to decide between a lottery that pays zero or 300 € with equal probabilities and a safe payment. The safe payment increased from 10 € to 200 € in steps of 10 €. The probability that one randomly selected decision would be implemented was 1/9. We use the standardized switching row from choosing the lottery to choosing the safe amount as our measure of willingness to take risk.

Mothers' *altruism* was elicited in the same way as for the children except for the fact that the mothers played for money and the anonymously assigned receiver was an unknown other participant. The mothers had to decide between 16 € for themselves and 4 € for the other participant (16,4) or 10 € for themselves and 10 € for the other participant (10,10). The mothers who chose (10,10) are labeled as altruistic. 84.0% percent of the mothers chose the altruistic distribution.

Panel B: Young adults

Sample. Data for the young adults feature the same preferences as for the preschool children and were collected in the BonnEconLab at the University of Bonn. 412 students took part in a series of experiments, 212 of them answered an additional take-home-survey in which they were requested to ask their parents how long they were breastfed. 175 students were breastfed.

Experimental measures. Time and risk preferences of students were elicited in a similar manner as for mothers in the preschool sample and as in Dohmen et al. (2010). To measure the students' *time preferences* they were faced with trade-offs between a smaller but sooner available reward and increasing larger but delayed rewards. The smaller sooner reward was fixed to 1600 points and the larger later reward also started at 1600 points and was increased 24 times by 2.5% p.a. assuming semi-annual compounding (100 points correspond to 0.8 €). To reduce measurement error, students played four different versions of this experiment in a random order. In the first version the sooner payment date was "today" and the later in six months. In the second version the sooner payment date was also "today" but the later was in 12 months. In the third version the payment dates were in six and in 12 months and the fourth version was a perturbation of the second version. Participants were informed that one decision would be randomly selected and paid. They also knew that the money was sent by mail irrespective of the payment date. We take the average first switching row from the sooner to the later payment as our measure of the students' time preference. For a more intuitive comparison with the results of the children the switching row was mirrored such that small values indicate impatience and high values indicate patience. To measure *risk preferences* of students they played two versions of an experiment where they had to decide between a lottery that pays zero or 1000 points with equal probability, and a successively increasing safe payment. The safe payment increased in steps of 50 points from zero to 1000. The two versions were played in random order and differed only in the exact size of the increase: In one version the increase in safe payments was in steps of exactly 50 points while in the other the increase was 50 points +/- 10 percent, i.e., with slight perturbation. One decision from both experiments was paid. 100 points corresponded to 0.8 €. We calculated the average first switching point from lottery to safe payment as our measure of the willingness to take risks. To measure *altruism*, we elicited responses to a three-item questionnaire. For the children the receiver could be a classmate (but not indicated who exactly) or a child from another unknown class. To match this situation for the young adults, we asked the question “How would you assess your willingness to share with others without expecting anything in return, concerning the following groups . . . ?” combined with the items “people from my neighborhood”, “people from my city” and “strangers”. Each item was answered on an 11-point Likert scale. We aggregated the survey answers using a principal component analysis.

Details on historical variation in breastfeeding duration

Research on time use (Sayer et al., 2004; Gauthier et al., 2004) and recent work in sociology (Hays, 1996; Faircloth, 2014) suggest that the pronounced variation in breastfeeding duration from the 1970s until the 1990s in Germany reflects general improvements in early life circumstances of children in this period through channels such as enhanced maternity leave legislation, and a general trend towards intensive parenting (Heimerdinger, 2009). To complement the evidence based on our cross-section analyses we therefore investigate whether the heterogeneity in breastfeeding durations across birth cohorts is related to heterogeneities in preferences. To test this conjecture we combined average historical breastfeeding data with preference measures from the German Socio-Economic Panel. We show that the pattern between early life circumstances and preferences found in our two cross-section data sets is also present in a panel based cohort analysis. Due to lack of administrative data we use breastfeeding data recorded for the purpose of marketing research by Nestlé. The information on breastfeeding quotas and durations are based on 250 interviews (per wave) of mothers of newborns. The interviews were conducted every second year starting in 1976. Figure A3.2 shows the data and reflects the historical variation in breastfeeding. Although these data were not conducted for scientific purposes they fit the pattern of fragmental information concerning this time (see the study of Heimerdinger (2009) and references therein).

We combine these data on breastfeeding with preference related data of the German Socio-Economic Panel (SOEP, 2012; Wagner et al., 2007) which is a representative panel survey of private households and persons in Germany. It contains about 11,000 households and more than 21,000 individual respondents. Since 2003 validated preference related questions have been part of the survey (Becker et al., 2012).

As a measure for time preference we use answers to the following survey question: "How would you describe yourself: Are you generally an impatient person, or someone who always shows great patience?". Participants gave an answer on an 11-point Likert scale where zero means "very impatient" and ten means "very patient". This time preference question was part of the survey only in 2008 and was validated with respect to incentivized experiments (Vischer et al., 2013). Concerning risk preference we use the answer to the question: "How do you see yourself: Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?". Answers were also given on an 11-point Likert scale where zero means "risk averse" and ten means "fully prepared to take risks". This question was asked in the six waves 2004, 2006, 2008, 2009, 2010 and 2011. The general risk question has been studied in various papers and was validated using incentivized experiments in representative samples as well as using behavioral evidence (Dohmen et al., 2011). Altruism towards other people was measured by the question how important it is for the participant "to be there for others". Answers were given on a 4-point scale. The altruism question was asked in waves 2004 and 2008 (Becker et al., 2012).

In our analysis we match the shares and the mean durations of breastfeeding (given initially breastfeeding) with preference measures from the SOEP by year of birth. Since the available breastfeeding data does not include citizens from the area of the former German Democratic Republic we also exclude these subjects from the preference data. Due to the lack of individual breastfeeding information we cannot exclude the not breastfed subjects as we did in the cross-sectional analysis. Therefore we adapt our estimation strategy such that we include the share of breastfed individuals and run the following aggregate level pooled OLS estimation:

$$\text{Pref}_{yt} = \beta_0 + \beta_1 \left[P(\text{ever BF}_y) \times \text{BF Duration}_y\right] + \beta_2 \left[1 - P(\text{ever BF}_y)\right] + \beta_3 \text{Age}_{yt} + e_{yt}$$

where Pref_{yt} indicates the average preference measure of individuals born in year y, measured in year t, $P(\text{ever BF}_y)$ is the share of individuals born in year y who were ever breastfed, BF Duration_y is the average duration of breastfeeding of individuals born in year y (given they were initially breastfed) and Age_{yt} is the Age of birth cohort y measured in year t. Therefore the marginal effect of an increase in

BF Duration is $\beta_1 \times P(\text{ever BF}_y)$ and indicates the effect of an increase in breastfeeding duration weighted by the share of breastfed children. To illustrate, β_1 shows the effect of an increase in BF Duration if all children would be initially breastfed.

Given the information concerning breastfeeding for a given birth cohort for every second year starting in 1976 and the preference measures in the SOEP we construct a panel on the cohort level including repeated measures of risk and altruistic preferences and a cross-sectional data set concerning time preference. Since the participants in the main part of the SOEP are interviewed for the first time at the age of 18 we can match birth-year-averages of breastfeeding durations to every cohort which was at least 18 years old at the date of a given wave. As the time preference question was only asked in 2008 we only yield eight observations in this case (the birth cohorts 1976, 1978, 1980, 1982, 1984, 1986, 1988 and 1990). Concerning risk preference we have observations based on six birth cohorts in wave 2004, followed by seven in 2006, eight in 2008, eight in 2009, nine in 2010 and nine in 2011. This yields 47 pooled observations. For altruism we got six birth cohorts in 2004 and eight in 2008, resulting in 14 pooled observations. The average preference measures are based on 116 to 254 individuals per wave and birth cohort.

We perform a pooled OLS estimation on risk and altruistic preferences and cluster the standard errors by birth cohort. Due to the lack of repeated measures concerning time preference we perform a cross-section OLS analysis. Since age is varying across waves the panel structure of our data concerning risk and altruism enables us to disentangle the breastfeeding duration effect from the age effect. Table A3.7 presents the results and indicates that the pattern found on the individual level is also present in a cohort level panel data analysis: breastfeeding duration is again negatively associated with willingness to take risk ($p < 0.05$) and positively associated with altruism ($p < 0.01$) and patience (not significant).

Chapter 4

Formation of human prosociality: Causal evidence on the role of social environment

4.1 Introduction

Prosociality is a particularly important aspect of human personality and affects a wide range of social and economic outcomes such as the provision of public goods, contract enforcement, charity, management of commons, financial development, governmental and judicial efficiency, redistribution and economic growth (Arrow, 1972; Knack and Keefer, 1997; Zak and Knack, 2001; Ostrom et al., 2002; Fehr et al., 1997; Fehr and Gächter, 2002; La Porta et al., 1997; Guiso et al., 2009). Despite its fundamental importance for the well-being of individuals and societies, little is known about how human prosociality forms, in particular about the causal effect of social environment[1]. This is not surprising, as it requires random assignment of life circumstances, and valid instruments to measure prosociality. We address these challenges by implementing a random variation of the environment and by measuring prosociality using different sources and established measures.

Our research strategy builds on the conceptual framework suggested by Cunha and Heckman (Heckman, 2006; Cunha and Heckman, 2007, 2008). They highlight early childhood as the critical and sensitive period in the human development

[1]For descriptive evidence on development patterns Fehr et al. (2008); Almås et al. (2010); Sutter and Kocher (2007); Fehr et al. (2013)

process. Accordingly, our sample consists of elementary school children. Moreover, Cunha and Heckman identify two primary channels responsible in the skill formation process, parental background and investments[2].

Here, we study both channels. In terms of parental background we compare children with different socio-economic status (high vs. low). In addition, we investigate the role of intergenerational transmission of prosociality, i.e., the correlation of prosociality between children and their mothers. To study the role of investments we randomly assigned children to an enriched social environment in the form of a mentoring program.

Our sample consists of 607 primary school children (47.0% are girls, age at the start of the program: mean= 7.76 years, std. dev.= 0.48) and their mothers who were recruited using official registry data (see Appendix A4 for details). Families were informed via postal mail about the possibility to take part in a study on child development and potentially a mentoring program. Families interested in participating had to send back a short questionnaire on socio-economic characteristics of the household and to state their willingness to have their child participate in the mentoring program and the interview. Using this information a household was classified as low SES if at least one of the following three criteria was met: low income, low education or single parent[3]. All other households were classified as high SES. We invited all low SES families and a randomly chosen subset of high SES families to participate, yielding 113 high SES and 494 low SES households (for details of the recruiting procedure see Appendix A4).

In total we thus study outcomes of three distinct groups. From the 494 low SES households 180 children were selected to participate in the intervention, using stratified random sampling[4]. This constitutes our intention to treat group (Treatment

[2]Their research builds on a dynamic model of skill formation where the technology of skill production is denoted by $\theta_{t+1} = f_t(h, \theta_t, I_t)$. θ_t stands for the vector of skill stocks at time t, h stands for parental characteristics such as personality and SES and I_t stands for the investments in children in at time t.

[3]Low income: equivalence income of the household is lower than 1.065 Euro, which corresponds to the 30% quantile of the German income distribution); Low education: both mother and father of the child have at most secondary education, i.e., are not qualified for university studies; Single parent households: Single parent who is not living together with a partner.

[4]Stratification included the three qualification criteria for low SES and place of residence (Cologne

Low SES)[5]. The remaining 314 children with low SES background form our intervention control group (Control Low SES). The third group consists of 113 children with high SES background (Control High SES). To study effects of parental background we compare prosociality of untreated children, i.e., Control Low SES and Control High SES. To investigate the effect of the investment in the form of a mentoring program, we compare outcomes between Treatment Low SES and Control Low SES.

4.2 Intervention, measures of prosociality and empirical strategy

The intervention we randomly implemented was a pre-existing non-profit mentoring program (see Appendix A4 for details). In this program, children are provided with a mentor for the duration of one year. Mentors are mainly university students (age at the start of the program: Mean = 23.76 years, std. dev. = 2.63) who signed up as volunteers. Conceptually, the idea of the program is to extend a child's horizon through engaging in joint activities with a new contact or attachment person, enhancing experiences and skills which are potentially scarce in the given family context. The most important feature of the program is the mere existence of a person responsive to a child's individual needs, strengths, weaknesses, and interests. The child experiences that an unrelated person spends time and effort, and takes care and responsibility for someone else. The program is meant to foster the acquisition of new skills on a purely informal basis. The informal character of the program therefore differentiates it from interventions meant to increase formal achievements such as educational attainment[6]. On a practical level, children met with their mentor once a week and engaged in versatile activities such as visiting the zoo, museum, or playground, reading, cooking, ice skating or just having a conversation.

---

and Bonn), the selection was random conditional on place of residence, see discussion and Appendix A4 for details.

[5]133 of these children were actually matched with a mentor. The reason for not matching all ITT children was the unexpected lack of voluntary mentors. We define a child as matched to a mentor if he had at least one meeting with the mentor.

[6]For an overview regarding different types of interventions and their effects see Rodríguez-Planas (2012); Heckman and Kautz (2014)

Before and after the intervention, children and their mothers[7] from all three groups, Treatment Low SES, Control Low SES, as well as Control High SES, were interviewed by trained interviewers (see Appendix A4 for details). Children participated in incentivized choice experiments and answered a short questionnaire. During the time her child participated in the experiments, the mother filled out a extensive questionnaire covering socio-economic background information and assessments of personality and attitudes regarding her child and herself.

To measure a child's expressions of prosociality in a comprehensive manner, we elicited three facets: altruism, trust and other-regarding behavior in everyday life (for detailed descriptions and protocols see Appendix A4). We elicited altruism using three incentivized choice experiments, one simple binary choice game as well as two continuous dictator games. Running variations of a similar game generates multiple measures of a child's altruistic behavior and thus allows reducing measurement error. Importantly, decisions in the experimental games had real consequences. We implemented an experimental currency called "stars". At the end of the experiments, children could exchange the collected amount of stars into toys. These were arranged in four categories, which visibly increased in objective value and subjective attractiveness to children. During the experiments, children knew that more stars would result in the option to choose toys from a higher category.

Following the procedures of Fehr et al. (2008) in the binary choice game, a child had to decide between two possible allocations of two stars, between himself and another unknown child of similar age and from the same city. In one allocation the decision maker received two stars, while the other child received zero stars (2,0). In the alternative allocation both decision maker and recipient received one star each (1,1). Both possible allocations were physically shown to the children and interviewers checked whether the children had fully understood the implications of each allocation. We also ran two so-called dictator games. In both versions of this game, interviewers showed the children two paper bags, one belonging to the

[7] Actually, 95% of the children were accompanied by their biological mother, 3% by their biological father, 3 children by a step or foster parent and we do not have unambiguous information on the accompanying person for about 2% of the children. We use the term "mother" for the adult accompanying the child.

interviewed child and the other belonging to another child, the receiver. Between games we varied the receiver. In one game the receiver is a child living in a city nearby. In the other game the child lives in an African country. Subjects knew that the African child does not live together with his parents since the latter are either “ill or dead”. In both versions, children were endowed with six stars and could choose how to distribute the six stars between the two bags. Our joint measure of altruism is the average share a child gives away in the three experiments[8].

Trust was elicited as part of a survey. Children had to state how much they agree to three statements on a five-point Likert scale, ranging from “totally correct” to “totally incorrect”. The scale was printed on an extra paper sheet. The interviewer explained the procedure and scale use with a simple neutral example item (“I like Spaghetti”). The trust items are age-adapted questions on the basis of the trust questions used in the German Socio Economic Panel Study (SOEP)[9] and 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 average rating on the Likert scale over the three items is our measure for a child’s trust[10].

Finally, we elicited other-regarding behavior of the child using survey responses of the mothers. As part of the survey, mothers answered the Strength and Difficulties Questionnaire (SDQ), which is a well-established behavioral screening questionnaire (Goodman, 1997). Statements about their child were rated on a seven point Likert scale, ranging from “does not apply at all” to “applies completely”. In the analysis we focus on the seven items which refer to a child’s other-regarding behavior, which read as “My child. . . ” “Shares readily with other children (treats, toy, pencils etc.)”, “Is helpful if someone is hurt, upset or feeling ill”, “Often fights with other children or bullies them” (reversed), “Gets along better with adults than with other children” (reversed), “Is generally liked by other children”, “Is kind to younger children”,

[8]Mean $= 0.390$, std. dev. $= 0.156, N = 604$; due to failed control questions or missing answers three children were excluded from the analysis of altruism

[9]These questions were experimentally validated in Fehr et al. (2002) and significantly predict choices in trust game experiments.

[10]Mean $= 3.193$, std. dev. $= 0.765, N = 607$; all 607 children answered all three trust question.

"Often volunteers to help others (parents, teachers or other children)". The average rating on the Likert scale over the seven items is our measure of other-regarding behavior in everyday life of the child[11].

In sum, we obtained three facets of prosociality, altruism, trust and other-regarding behavior, using incentivized choice experiments, experimentally validated survey responses and well-established survey items elicited from the mother. Thus our measures capture a broad characterization of prosocial disposition based on different elicitation methods, and statements from different sources (mother and child). In the analysis below we use standardized measures (z-scores, mean = 0, std. dev. = 1) in order to enhance comparability. We also collapse the three individual measures into one joint measure of prosociality, interpreting altruism, trust and other-regarding behavior as facets of an underlying prosocial disposition. To detect this joint underlying trait (and further reduces measurement error) the joint measure is constructed using principal component analysis[12]. Using responses to validated survey items we also generate a measure of prosociality for mothers and mentors (see Appendix A4 for details). As for children, the measure consists of the three facets altruism, trust and other-regarding behavior.

Concerning the effect of parental background, following recent findings (see Bauer et al. (2014) and Chapter 3 herein), we hypothesized that additional material and cognitive resources available in high SES households have a positive effect on prosociality relative to low SES households. In light of recent empirical and theoretical work[13] we further expected a positive impact of mothers' prosocial attitudes, i.e., an intergenerational transmission of prosociality. Concerning the effect of the intervention, evidence on social learning (Bandura, 1965, 1986) suggests that exposure

[11]Mean = 5.885, std. dev. = $0.811, N = 604$; to limit missing values due to incomplete mother questionnaires, we include an observation if at least four out of seven items are completed; three missing values remain.

[12]We performed a principal component analysis using the standardized measures of altruism, trust and other-regarding behavior, resulting in one component according to the Kaiser Criterion (Eigenvalue > 1).

[13]For an example of theoretical work see Bisin and Verdier (2001). Empirically, for one of our three facets, trust, Dohmen et al. (2012) have in fact shown a strong and systematic intergenerational correlation. Similarly, risk attitudes and time preferences are positively correlated between parents and children, see Dohmen et al. (2012) and Kosse and Pfeiffer (2012).

to role models affects prosocial behavior. The investment under study provides the child with such a positive role model. The mentor represents an additional attachment person, responsive to a child's individual needs, who spends time and effort, and takes care and responsibility for an unrelated person (the child). Mentors display a relatively high level of prosociality (see below). We therefore hypothesized a positive effect of the mentoring program on the prosociality of children, in particular in those households where prosociality is rather low.

In presenting our results we first analyze the impact of parental background. We then show our main finding, the treatment effect of the intervention on prosocial dispositions in children. Throughout, we report intention-to-treat (ITT) estimates[14]. The measures of prosociality that we use in the main analyses were elicited in the second interview that took place right after the intervention period. Below we also use measures from the first interview (before assignment into the mentoring program), to address potential concerns such as baseline balance and attrition bias, and also refer to alternative estimates.

4.3 Parental background: SES and mothers' prosociality

Fig. 4.1 (left panel) displays a pronounced SES effect. For the two untreated groups of children, Control Low SES and Control High SES, the figure shows that children from high SES households score significantly higher on the prosociality measure than children from low SES households ($p < 0.01, N = 424$, two sided t-test). In terms of effect size, the difference amounts to 35.4% of a standard deviation. The right panel of Fig. 4.1 additionally shows that mothers' prosocial attitudes have a strong impact on children's prosociality. It compares the levels of prosociality for children whose mothers either score above or below median on the prosociality measure ($p < 0.01, N = 417$, two sided t-test). The intergenerational correlation of prosociality is 0.228 ($p < 0.01, N = 417$, Pearson correlation coefficient, Control Low SES and Control High SES).

[14]An ITT analysis is based on the initial treatment assignment and not on the treatment eventually received

These results reveal that parental background significantly affects formation of prosociality, both in terms of SES and intergenerational transmission. Importantly, we observe a gap in the development of prosocial attitudes for elementary school children in response to different socio-economic environments. This sets the stage for analyzing the potential effect of the intervention on low SES children.

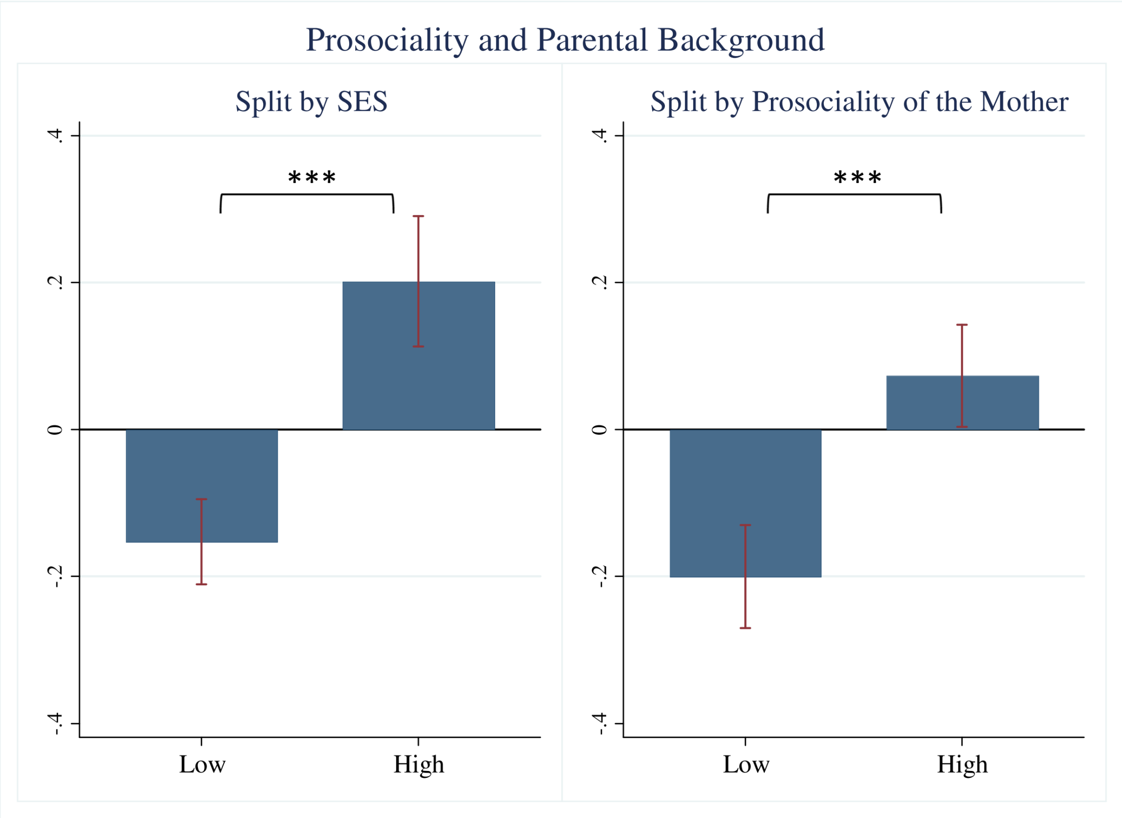


Figure 4.1: Prosociality and Parental Background. The left panel shows higher levels of children's prosociality in the Control High SES group compared to the Control Low SES group. The right panel shows higher levels of prosociality for children of highly prosocial mothers (median split). The prosociality measures of mother and child are constructed using a principal component analysis for the aggregation of altruism, trust and other-regarding behavior, respectively. The scale on the y-axis indicates z-scores (i.e. standardized measures) of children's prosociality. Error bars show standard errors of the means (SEM). *** indicates significant differences at the 1% level (two-sample *t*-tests, N(left panel) = 424, N(right panel) = 417, difference due to incomplete mother questionnaires).

4.4 Investment: The causal effect of social environment on prosociality

Figs. 4.2 and 4.3 show our main results. Fig. 4.2 displays a positive and significant treatment effect for the joint measure of prosociality. Children whose social environment was randomly enriched through participation in the mentoring intervention score 29.5% of a standard deviation higher on the prosociality measure than children from the control group ($p < 0.01, n = 489$, two-sided t-test). Fig. 4.2 also shows that the high-low SES developmental gap shown in Fig. 4.1 is substantially reduced for children who participated in the mentoring program. In fact children from Low SES Treatment and High SES Control score very similarly on the prosociality measure. The difference between high and low SES children is no longer statistically significant ($p = 0.605, n = 289$, two-sided t-test).

The positive effect of enriching the environment is not only sizeable and significant using the joint measure, but also when looking at three measures independently (see Fig. 4.3). Children in the Low SES Treatment group are more altruistic, reveal higher levels of trust and show more prosocial other-regarding behavior than children in Low SES Control. In terms of altruism, children in Low SES Treatment give 18.6% of a standard deviation more compared to children in Low SES Control ($p < 0.05, N = 492$, two-sided t-test). Likewise, treatment children report significantly higher levels of trust with an increase of 20.9% of a standard deviation compared to the control group ($p < 0.05, N = 494$, two-sided t-test). Finally, mothers of treated children report more pronounced other-regarding behavior of their children in comparison to mothers from the control group. The difference amounts to 19.4% of a standard deviation ($p < 0.05, N = 491$, two-sided t-test).

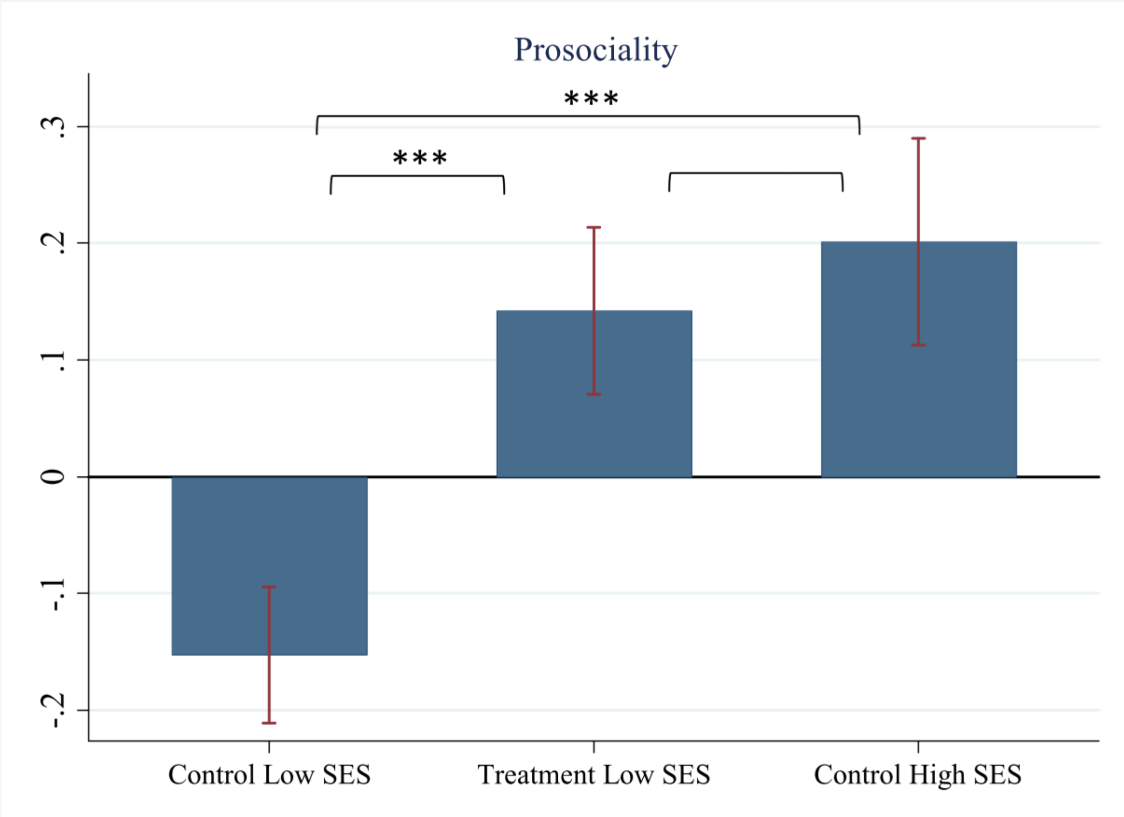


Figure 4.2: Significantly higher levels of prosociality for treated children compared to untreated children (Treatment Low SES vs. Control Low SES). Reduction of developmental gap: There is no significant difference between Treatment Low SES and Control High SES group. The prosociality measure is constructed using a principal component analysis for the aggregation of altruism, trust and other-regarding behavior. The scale on the y-axis indicates z-scores (i.e. standardized measures) of children's prosociality. Error bars show standard errors of the means (SEM). *** indicates significant differences at the 1% level (two-sample t-tests, N(Treatment Low vs. Control Low) = 489, N(Control High vs. Control Low) = 424, N(Treatment Low vs. Control High) = 289).

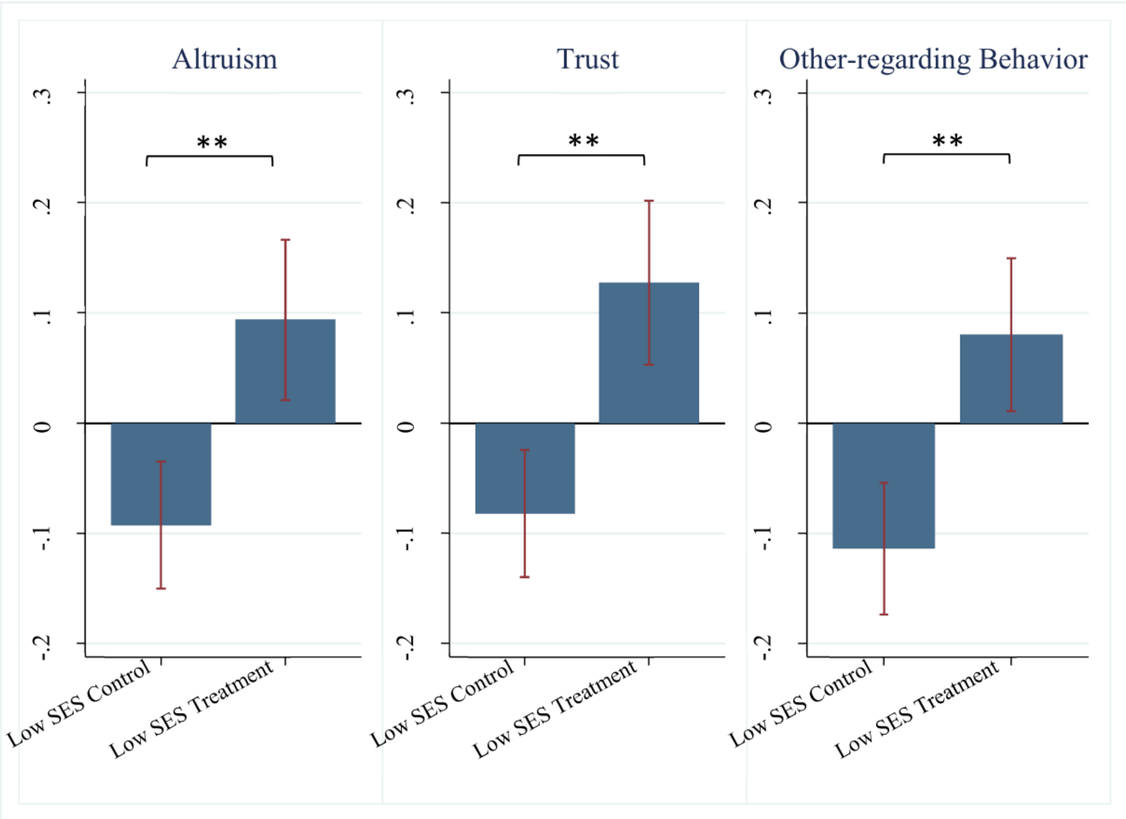


Figure 4.3: Significantly higher levels of altruism, trust and other-regarding behavior for treated children compared to untreated children (Treatment Low SES vs. Control Low SES). Altruism is measured in three different incentivized dictator games. The trust measure is based on children's responses to a questionnaire. Other-regarding behavior is based on a questionnaire rating by the mother with respect to other-regarding behaviors of their child in every-day life. The scale on the y-axis indicates z-scores (i.e. standardized measures) of respective prosociality facet of the child. Error bars show standard errors of the means (SEM). ** indicates significant differences at the 5% level (two-sample *t*-tests, N(altruism) = 492, N(trust) = 494, N(other-regarding behavior) = 491. Number of observations varies due to failed control questions or missing answers).

4.5 Discussion

In sum, the results show that both parental background and investments affect children's prosociality. We now combine both perspectives to answer the question whether the investment has differential effects for children with different parental background. To answer this question we ran OLS regressions where prosociality of children (Low SES Treatment and Low SES Control) is regressed on a treatment dummy, parental background variables (prosociality of mothers, income, education and single parent status), and the respective interactions (see Table A4.1). The analysis reveals that children whose mothers score low on our prosociality measure benefit most. The respective interaction coefficient is significant ($p < 0.1, N = 479$) and indicates that the treatment effect increases by 15.6% of a standard deviation of prosociality if the mother scores one standard deviation lower in prosociality. A similar reinforced effect is found for children whose parents have relatively low levels of education, but the respective interaction effect is slightly above the 10-percent significance level ($p = 0.131, N = 489$, column (2)). The interaction effect with income is very small and insignificant; the one for single parent status is larger but insignificant as well. These findings suggest that the mentoring treatment particularly benefits children whose mothers display low levels of prosociality, and who come from relatively low educated households. Mentors provide exactly these two resources. First, nearly all mentors are either university students or hold a university degree[15]. Second, they score very high on our prosociality measure. Compared to the mothers they score 39.7% of a standard deviation higher in prosociality ($p < 0.01, N = 692$, two-sided t-test)[16]. Taken these findings together, this suggests that the mentoring program benefits children by providing resources and stimuli that are scarce at the household level.

[15] 99 mentors answered our questionnaire. 97 mentors provide information on their education. All of them hold at least a university entrance diploma, 93 of them are students or have already gained a university degree.

[16] It is not surprising that mentors display a particularly high prosocial motivation given that they are self-selected volunteers who spend a considerable amount of their time to serve in a mentoring program devoted to support children, i.e., they engage in an altruistic activity at a personal cost.

4.6 Robustness checks and alternative estimates

Potential confounds in randomized controlled studies concern baseline balance and attrition bias. The fact that we have elicited all reported measures not only after but also before treatment assignment allows us to directly address these potential concerns. The analysis confirms that the randomization procedure was successful since the pre-treatment measures of prosociality do not differ by treatment status (see Table A4.2). The lost to follow-up rates are generally low (16.3% among all Low SES) and do not differ between treatment and control group ($p = 0.563, N = 590$, two-sided test of proportions). Moreover, pre-treatment measures of prosociality, treatment assignment and their interactions are not related to participation in the post-treatment interview. This indicates the absence of any outcome-related systematic attrition (see Table A4.3), and confirms the validity of the presented results. Since our experimental design used conditional random assignment (conditional on the place of residence of the families, see Appendix A4 for details), we also present estimations using city fixed effects (see Table A4.4) for our main results (displayed in Fig. 4.2 and Fig. 4.3). The estimates show that (facets of) prosociality do not differ by city and that controlling for city does not affect the estimated treatment effect.

We also provide treatment on the treated (TOT) estimates. While the presented ITT effects build directly on the random assignment, they understate the effect of actually receiving the treatment because not all children selected into the treatment group were actually matched with a mentor. Estimating treatment on the treated (TOT) effects by using random assignment as an instrument for actual treatment adds the assumption that being assigned to treatment had no effect on those who remained untreated. This assumption plausibly holds in our case since the vast majority of these families were not even informed about being in the ITT group[17]. The TOT effects provide policy relevant information as they reveal the expected effects of implementing a mentor-mentee match. Given a matching rate of 73.9% the TOT estimates exceed the ITT effects by about 35.3% (see Table A4.5).

[17]Due to an unexpected lack of mentors not all families in the ITT group were contacted and offered a mentor

Another method for the evaluation of randomized controlled studies, particular popular in the context of development economics, is to regress post-treatment measures of interest on a treatment dummy and the pre-treatment measure. This method outperforms post-treatment mean comparisons and differences-in-differences approaches regarding statistical power (McKenzie, 2012)[18]. Results using this method, which further confirm our results, are displayed in Table A4.6.

4.7 Conclusion

Our results indicate that there is a development gap in prosocial disposition of children from Low SES and/or low prosocial families but enriching the social environment bears the potential to positively impact the formation of prosociality. Relative to the control group, children accompanied by a prosocial role model become more altruistic and show higher levels of trust and more prosocial other-regarding behavior. The high-low SES difference is no longer significant. Given that treatment assignment was random, our study thus provides causal evidence on the effects of social environment on prosocial attitudes. Therefore, our findings establish the importance of life circumstances on character formation, and provide support for models of cultural evolution and theories that do not rely exclusively on genetic explanations (Richerson and Boyd, 2008; Henrich et al., 2004). Our findings also add to the discussion on growing inequality and the intergenerational transmission of life-outcomes and socio-economic status (Piketty and Saez, 2014; Aizer and Currie, 2014; Haushofer and Fehr, 2014). We have shown that substitutive investments such as the mentoring program under study have the potential to substantially reduce personality developmental gaps arising from differences in socio-economic status.[19]

[18]For a discussion concerning the different assumptions of these estimators see Imbens and Wooldridge (2009)

[19]For recent evidence on the effects of childhood interventions on life outcomes as income or health see Campbell et al. (2014); Gertler et al. (2014).

Chapter 5

Unfair pay and health

5.1 Introduction

A large and growing body of evidence suggests that fairness perceptions play an important role in labor relations, affecting work morale, effort provision and market efficiency (see e.g., Fehr et al., 1993, 1997; Abeler et al., 2010; Charness and Kuhn, 2011; Kube et al., 2012; Altmann et al., 2014).[1] Fairness considerations have also been shown to help reconciling evidence on non-standard effects of minimum wages (Falk et al., 2006; Katz and Krueger, 1992; Card, 1995). While this work has studied behavioral effects of fairness perceptions, the present paper provides evidence on adverse effects of unfair pay on the physiological level. In particular, we investigate the potential impact of unfair pay on stress and adverse health outcomes. To test for the potential link between wage related fairness perceptions, stress and health, we use an integrated approach, combining lab and field data to exploit complementarities of different data sources. We proceed in two steps. First, we report results from a lab experiment testing the hypothesis that unfairness perceptions have a negative effect on heart rate variability (HRV). A low HRV is a stress related early indicator of functional and structural impairments of the cardiovascular system, which increases the probability of future manifest coronary heart disease (see e.g., (Steptoe and Marmot, 2002; Dekker et al., 2000; Gianaros et al., 2005)). Second, we analyze data

[1] For an overview and related studies, see (Fehr and Gächter, 2000). The above-cited experimental work is complemented by interview studies with personnel managers (see, e.g., Agell and Lundborg, 1995; Bewley, 1999, 2005). Akerlof (1982) provides an early theoretical analysis of fairness and labor market efficiency.

from a large representative data set to study whether our findings from the lab extend to the general population and the real-life labor market, in the sense that perception of unfair pay is related to (specific) health outcomes.

The lab experiment implements a simple principal-agent relationship. In the experiment an agent produces revenue by working on a tedious task. The principal receives the revenue produced by the agent and decides how to allocate it between the agent and himself. This set-up potentially generates various degrees of unfair pay, where the source of variation is the heterogeneity in generosity of the principals, who are randomly assigned to agents. Agents' HRV is monitored throughout the experiment. The experimental set-up allows us to precisely measure physiological responses, actual payments and revenues as well as agents' fairness perceptions of pay. Our hypothesis to be tested is an inverse relationship between the degree of unfair pay and HRV[2]. The results support this hypothesis. Perceptions of unfair pay are inversely related to agents' HRV, our measure of impaired cardiac autonomic control.

Building on our controlled laboratory evidence and the significance of HRV as long-run indicator for stress-related cardiovascular health, we further investigate whether perceptions of unfair pay are negatively correlated to health status in the general population. We test this hypothesis using data from the German Socio Economic Panel (SOEP), a large data set that is representative for the adult German population (Schupp and Wagner, 2002). In particular, we regress employees' subjective health status on whether they consider their wage as fair or unfair. Controlling for a large set of individual as well as labor market characteristics such as net wages, labor market status, occupational status, firm size and industry, we find a strong and significant association between perceptions of unfair pay and lower subjective general health status. We also perform dynamic panel estimations and find evidence for a Granger causal effect of unfair pay on general health. In light of our lab findings we further hypothesized that adverse health effects should be specific to diseases re-

[2]Note that low heart rate variability is observed, among others, during states of mental stress while enhanced heart rate variability occurs during states of mental relaxation (for details and references, see Section 5.2). This is why we expect an inverse relationship between unfairness and HRV.

lated to the nervous system and the experience of stress, such as heart disease and high blood pressure. Testing for an effect on specific health outcomes is possible as the SOEP not only elicits subjective responses to general health outcomes but also with respect to specific diseases. Confirming our hypothesis, we find that perceptions of unfair pay are in fact mainly correlated to cardiovascular health outcomes. No such relation is observed for diseases such as cancer or apoplectic stroke. The effects are most pronounced for full-time employees above age 50. This is what we would expect given that the visible occurrence of cardiovascular diseases usually does not start before age 50 (Roger et al., 2012), and experience of unfair pay (the stressor) is likely to be more affective the longer the working experience.

Our findings establish a link between unfair pay and coronary heart disease suggesting that on top of behavioral consequences reported in previous work, perceptions of unfair pay can have important negative physiological consequences with possible welfare implications: The global public health and economic burden of cardiovascular disease is immense. By the year 2020, coronary heart disease, along with major depression, is estimated to be the leading cause of life years lost to premature death and years lived with disability worldwide (Lopez et al., 2006). Moreover, among adult populations of high income countries, coronary heart disease is the leading cause of death, and cost of illness studies estimate that almost one percent of the gross national product is attributable to the direct and indirect costs of coronary heart disease (Liu et al., 2002). On an organizational level our findings suggest that fair pay does not only contribute to higher work moral and motivation, but also to a better health status of employees. In this sense our findings suggest important efficiency consequences of fair wages, additional to efficiency wage arguments (Akerlof, 1982).

The remainder of the paper is organized as follows. In the next section we present our experimental design and results. Section 5.3 reports results regarding the representative sample, including cross-sectional and dynamic panel estimations. Section 5.4 concludes.

5.2 An experiment to study physiological responses to unfair pay

Experimental design and procedural details. In the experiment we implemented a simple principal-agent relationship. Upon arrival to the lab, subjects were randomly assigned to the role of agent or principal and randomly matched into pairs consisting of one agent and one principal. The interaction was completely anonymous, i.e., at no point subjects learned about the identity of their partner. Subjects received all instructions via computer screen.[3] We used z-Tree as computer software (Fischbacher, 2007). Agents received a pile of numbered sheets. On each sheet there was a table containing a large number of zeros and ones. The work task was to count the correct number of zeros on a given sheet and to enter this number on a computer screen. Total working time was 25 minutes. Each correctly entered number of zeros per sheet created revenue of three Euros. If the entered number was "almost" correct (deviation of plus/minus 1 with respect to the correct number) revenue was one Euro. The accumulated revenue was continuously shown to agents on the screen. Agents were explicitly told that they could complete as many sheets as they wanted to, including completing no sheet at all. Principals were informed that agents created revenue by working on a task. They did not work and were told that they were free to do things like reading newspapers, completing class-work etc.

After completion of the 25 minute working time, each principal was informed about the accumulated revenue created by his agent and was asked to allocate it between himself and the agent. Before the principal's allocation decision was communicated to the agent, the latter was asked to state the amount of money he would consider to be an "appropriate pay". This information was not revealed to the principal. The agent was then informed about the principal's actual allocation decision. Starting with this feedback, the agent was given a time window of 15 minutes to cope with this information.[4] During this time subjects answered a short survey on perceived fairness of the received payment. We used the following item (Fairness

[3]Instructions are shown in Appendix A5.

[4]This is a standard procedure in HRV studies. Brosschot and Thayer (2003) show that especially negative emotions are related to a relatively long lasting heart rate response.

question): "In your view, how fair was the return you received from your principal?" Answers were given on a 5-point Likert scale, with higher values indicating that returns were considered less fair.

As physiological measure of agents' autonomic nervous system activity we used heart rate variability (HRV), an established indicator of stress-related activation of the autonomic nervous system (Task Force, 1996; Steptoe and Marmot, 2002)[5]. HRV reflects the continuous interaction of sympathetic and vagal influence on heart rate, indicating an individual's capacity to generate regulated physiological responses to demanding situations (Appelhans and Luecken, 2006). Low HRV mirrors a decreased vagal tone with sympathetic predominance and is observed, among others, during states of mental stress (von Borell et al., 2007). Conversely, enhanced HRV occurs during states of mental relaxation (Vermunt and Steensma, 2003). A low HRV is an early indicator of functional and structural impairments of the cardiovascular system, which increases the probability of future manifest coronary heart disease (Steptoe and Marmot, 2002; Dekker et al., 2000; Gianaros et al., 2005). In the analysis we use two measures of HRV. The first one serves as a baseline measure (HRV_baseline) and was measured towards the end of the working period but prior to the revelation of the allocation decision. The second one was taken 15 minutes after exposure to the stimulus, i.e., the revelation of the principal's allocation decision. It records the response of the autonomic nervous system to the stimulus and

[5]At the beginning of the experiment a polar F810i device (polar electro OY, Kempele, Finland) was attached to record and store time intervals between consecutive heart beats (inter-beat-interval, IBI). Agents were instructed to remain seated during the whole experiment and try to restrict all movements, with the exception of their dominant arm operating the computer. The target time window for physiological recordings lasted five minutes. Data were transmitted to a PC, stored, and analyzed offline by a researcher who was blind to the psychological outcome measures. After visualizing and manually correcting data for artefacts a smoothness priors method was used to remove trends of the IBI time series. Then, a HR time series was derived and the following time-domain based HRV indices were calculated: SD-IBI (standard deviation of the IBI series), SD-HR (standard deviation of the HR series), and RMSSD-IBI (root mean square of successive differences of the IBI series) (Niskanen et al., 2004). The RMSSD-IBI represents a sensitive index of parasympathetically-dominated, respiratory related, fast fluctuations of HR, and can be calculated with milliseconds precision. It is considered to accurately index resting vagal tone directed to the heart and was documented to be rather resistant to the biasing effects of breathing (Penttilae et al., 2001). As SD-IBI and SD-HR are highly correlated with RMSSD-IBI we restrict the presentation of findings to RMSSD-IBI, as a robust and well validated time-domain based indicator of parasympathetic cardiac control. All calculations were done with a computer program for advanced HRV analysis (Niskanen et al., 2004).

serves as dependent variable (HRV_response).

Subjects were male students from the University of Bonn studying various majors except economics. They gave their informed consent to participate in the experiment. Exclusion criteria were the use of medication with potential interference with cardiovascular function or the presence of a chronic disease condition, such as hypertension, cardiac arrhythmias, coronary heart disease, or diabetes. In total 80 subjects participated in the experiment (40 principals and 40 agents). During the process of data collection, we had to exclude data of 10 subjects in the role of agents, due to incomplete heart rate measurements. The main analysis is thus based on 30 subjects in the role of agents with complete data. Importantly, the 10 subjects who were excluded due to incomplete heart rate measurements were not different to the other subjects, neither in terms of working behavior nor treatment by their principals (see Footnote 7).

Experimental results. In our analysis we use three measures of perceived unfairness, i.e., how unfair agents perceive their principals' allocation decisions. The first measure is simply the difference between a principal's and an agent's payoff. It is informed by fairness theories that model fairness comparisons in terms of deviations from an equitable share[6]. Note that this measure considers wage payments and resulting payoffs only, disregarding effort costs. We have to abstract from effort costs given that in a real effort experiment, effort costs are unknown to the experimenter. The second measure is the difference between the payoff an agent indicated as "appropriate payoff" prior to knowing the actual allocation decision, and the actually received payoff. This measure therefore includes a subjective component of the agent and accounts for fairness perceptions that include both, payoffs as well as effort costs. The third measure concerns answers to the Fairness question, i.e., agents' assessments of how fair they perceived the wage payment of their principals (on a 5-point Likert scale). This measure completely abstracts from observed wage payments and allows for a fully subjective fairness assessment of agents. It is also similar to the survey measure we use in our analysis of the effects of fairness percep-

[6] See, e.g., Fehr and Schmidt (1999) or Falk and Fischbacher (2006) where fairness or unfairness is evaluated as difference in payoffs (equity as a reference standard).

tions on health outcomes in the general population. The three measures are highly correlated (Spearman's ρ is between 0.498 and 0.705, $p < 0.01$).

Table 5.1 reports means and standard deviations of our main variables[7]. On average agents produced total revenue of 20.93 Euro and indicated that they would consider a share of 14.03 Euro (67% of total revenue) as "appropriate payoff". This contrasts sharply with the amounts agents actually received. On average principals allocated 9.53 euros to agents (46% of total revenue).[8] Table 5.1 further shows the difference in payoffs of principals and agents, as well as the difference between the amounts considered as appropriate and the amounts actually received. Both differences vary considerably among subjects (standard deviations of 4.90 and 4.37, respectively). In other words the experiment generated substantial variation in (perceived) fairness violations, a prerequisite for the analysis of the effect of fairness perceptions on HRV.

Variable	**Mean**	**Standard Deviation**
Total revenue produced by agents (in Euro)	20.93	8.57
Payoff allocated to the principal (in Euro)	11.40	4.19
Payoff received by agent (in Euro)	9.53	5.58
Principal's - agent's payoff (in Euro)	1.87	4.90
Payoff seen as appropriate by the agent	14.03	6.68
Appropriate - actual payoff (in Euro)	4.50	4.37
Fairness question (scale: 1-5)	3.43	1.43

Table 5.1: Descriptive statistics. $N = 30$; appropriate refers to the amount, which is stated by the agent as appropriate pay after the total revenue was known but before the principal's allocation decision was communicated; the difference between principal's and agent's payoff is our first measure of unfairness, the second is the difference between appropriate and actual payoff and the third is the answer to the Fairness question; answers are given on a 5-point Likert scale and are coded such that higher values imply higher levels of unfairness.

To test our hypothesis of an inverse relationship between the degree of fairness violation and HRV we regress HRV_response on our three measures of unfairness.

[7] Table 5.1 reports data for the 30 subjects with complete heart rate measurement. Subjects with incomplete measurement were not different in any systematic way. Total revenue for this group was 20.20 (Std. dev. 7.23), the payoff allocated to the principal was 11.70 (Std. dev. 3.71), the amount received by the agent 8.50 (Std. dev. 5.23) and the amount seen as appropriate by the agent was 13.80 (Std. dev. 6.34). Kruskal-Wallis rank tests do not reject the null hypothesis that both groups are drawn from the same population (p-values are between 0.54 and 0.98).

[8] Only two agents received more than they indicated as an appropriate amount.

The results are shown in Table 5.2. To ease comparison, the measures of unfairness are standardized. All three coefficients are negative and significant, see columns (1), (3) and (5). These results indicate that HRV reacts negatively to perceptions of being treated in an unfair way, i.e., fairness systematically affects the autonomic nervous system. Columns (2), (4) and (6) include two important control variables, HRV_baseline and generated revenue. Controlling for different baseline levels addresses the possibility that subjects with a generally low baseline HRV have, e.g., systematically different fairness expectations or standards, and may therefore perceive payments differently. Likewise, it is important to control for levels of generated revenue to exclude the possibility that principals were willing to share relatively higher amounts with more productive agents. Results in columns (2), (4) and (6) show that our main result is robust to including these controls. While the coefficients of interest are slightly smaller compared to those reported in columns (1), (3) and (5), they remain significant.

	HRV_response					
	(1)	(2)	(3)	(4)	(5)	(6)
Principal's - agent's payoff	-5.361**	-4.717**				
	[1.960]	[1.976]				
Appropriate - actual payoff			-5.781***	-4.363**		
			[1.781]	[1.773]		
Fairness question					-6.514***	-5.724***
					[2.141]	[1.921]
HRV_baseline		0.457***		0.497***		0.491***
		[0.145]		[0.145]		[0.130]
Generated revenue		-0.451*		-0.207		-0.369*
		[0.232]		[0.222]		[0.181]
Constant	32.072***	30.483***	32.072***	24.408***	32.072***	27.927***
	[1.910]	[5.198]	[1.868]	[4.654]	[1.782]	[4.244]
Observations	30	30	30	30	30	30
R-squared	0.214	0.435	0.249	0.434	0.316	0.534

Table 5.2: Regression analysis on the relation between perceived fairness and HRV. OLS estimates with robust standard errors in brackets. ***, **, * indicate significance at 1-, 5-, and 10-percent level, respectively. The dependent variable is HRV_response, i.e., the heart rate variability, which was measured after exposure to actual payoff. It records the response of the autonomic nervous system to the stimulus. HRV_baseline measures the HRV towards the end of the working period. Appropriate refers to the amount, which is stated by the agent as appropriate pay after the total revenue was known and before the principal's allocation decision was communicated; the difference between principal's and agent's payoff is our first measure of unfairness, the second is the difference between appropriate and actual payoff and the third is the answer to the Fairness question; answers are given on a 5-point Likert scale and are coded such that higher values imply higher levels of unfairness. The unfairness measures are standardized (mean = 0, standard deviation = 1). Generated revenue represents total revenue produced by the agent.

5.3 Fairness perceptions and health: Representative field data

Our experimental data show that perceiving a wage as unfairly low induces impaired cardiac autonomic control. In view of the significance of HRV for stress related cardiovascular health, our results suggest potential effects on health outcomes as a reaction to perceptions of unfair exchange at work. In other words, we would expect that if perceptions of unfair pay constitute a chronic source of stress, unfair pay should be negatively related to employees' general health status and in particular to stress-related diseases. In the following we investigate this issue in the context of the German labor market by analyzing data from the German Socio-Economic Panel (SOEP). Exploiting complementarities between lab and field data is useful in terms of cross validating findings and simultaneously providing evidence that is both, controlled and based on representative data.[9]

The SOEP is a representative panel survey of the adult population living in Germany. All household members above age 17 are interviewed on a wide range of individual and household information and for their attitudes on assorted topics.[10] Each wave records information on the respondents' current labor market status, including wages. Due to data availability our main analysis is based on data of the year 2009 which also include an item regarding perceived fairness of wage payments.[11] The question reads as follows: "Do you consider the income that you get at your current job as fair?" with the possible answers "yes" or "no". Among the roughly 11,000 subjects, who are active in the labor market, about 36% stated that they consider their wage as unfair. The data set also contains items about health status, in particular about subjective health status in general and whether various diseases have been diagnosed in the past. The question about health status in general is: "How would you describe your current health status?" Responses were given on a 5-point scale ranging from "very good" to "bad". For the analysis the variable was coded in a way that higher values indicate better health. For the full sample the mean is 3.55 (standard deviation is 0.86). While subjective health indicators have their limitations, previous research in health economics suggests that responses to subjective health status questions predict labor market outcomes, health impairments and mortality.[12]

[9]For a discussion of lab and field data, see Falk and Heckman (2009).

[10]For more details on the SOEP, see www.diw.de/gsoep/ and Schupp and Wagner (2002), SOEP v28 is used.

[11]Not all items we use are elicited in every wave. Next to the Fairness question which was also asked in 2005, 2007 an 2011, in the year 2009 the questionnaire covers items about particular diseases and personality, which are essential for our analysis. The only exception is body mass index (BMI) which was not elicited in 2009. BMI data are therefore taken from the 2010 wave.

[12]For a comprehensive discussion of the literature, measurement issues, reporting biases and effects on labor market outcomes, see Currie and Madrian (1999). They discuss potential limitations of subjective health measures but also point out that self-reported measures are good indicators of health as they are highly correlated with medically determined health status. The authors thank Janet Currie for suggesting testing for selective associations.

A more "objective" measure can be constructed from answers to the question whether a physician has "ever diagnosed" a particular disease, mentioned in a list presented to participants. Analyzing responses to this question is particularly informative as it allows a more precise test of our hypothesis: Since impaired cardiac autonomic control is of particular significance for cardiovascular health, we hypothesized that perceptions of unfair pay predict stress-related diseases such as heart disease and high blood pressure, rather than diseases such as cancer or asthma. Finding selective associations would suggest that the main mechanism how fairness perceptions affect health operates through cardiac control similar to what we find in our lab data.

In Table 5.3 we report OLS estimates to assess how subjective health status is related to perceptions of unfair pay.[13] Since fairness perceptions may simply reflect relatively low wage levels we control for net wages. We also control for age and gender. Column (1) shows a negative, highly significant coefficient for unfair wage. Thus, respondents who consider their income as unfair report a significantly worse health status. Net wages and age have a significant effect on self-reported health status in the expected directions. Column (2) adds further controls, which may simultaneously affect fairness perceptions and health status, respectively. These include marital status, whether the respondent lives in East Germany, labor market experience (part and full time), educational background, firm size, occupational status (e.g., blue collar vs. white collar), type of industry and measures of personality. The complete specification and all coefficients are shown in Table A5.1 in Appendix A5. In column (3) of Table 5.3 we exclude employees for whom the relation between fairness perception and health status is less plausible. This includes employees who work only part-time and, in particular, the self-employed who largely determine their income themselves. Since visible occurrence of cardiovascular diseases usually does not start before age 50 (Roger et al., 2012), we additionally, in column (4), exclude employees who are younger than 50 years old.

Results in columns (2) and (3) indicate that the unfair wage coefficient is robust with respect to adding various controls and restricting the sample to full-time

[13]We get the similar results using Ordered Probit estimations.

employees. This means that conditional on wage level, educational background, labor market conditions, industry and labor market status, health status is strongly associated with the perception of receiving an unfairly low wage. As expected, the coefficient is somewhat larger in the specification that excludes part-time and self-employed workers. The fact that the coefficient of interest increases when moving from column (3) to column (4) further indicates that the observed negative relation between unfair pay and subjective health is more pronounced for the work force above age 50. Interestingly, an inspection of all coefficients in columns (2) to (4) of Table A5.1 (see Appendix A5) reveals that most control variables such as industry or firm size have no systematic effect on health status. The only systematic effect on top of unfair pay, net wages, gender and East German origin is found in respondents' personalities, measured with the Big-5 inventory[14]. The relevance of personality in this context is in line with Conti and Heckman (2010) who provide evidence for the importance of personality in determining health. Conscientiousness, extraversion and agreeableness are all positively related to better health conditions. Neuroticism, on the other hand, is negatively associated with health.

[14]The Big-5 can be broadly classified as follows: Openness to experience (appreciation for art, emotion, adventure, and unusual ideas; imaginative and curious), conscientiousness (a tendency to show self-discipline, act dutifully, and aim for achievement), extraversion (a tendency to seek stimulation and the company of others), agreeableness (a tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others), neuroticism (a tendency to easily experience unpleasant emotions such as anxiety, anger, or depression). See e.g. Almlund et al. (2011) and Becker et al. (2012) for an overview.

Dependent variable: subjective health status (higher values indicate better health)				
	(1)	(2)	(3)	(4)
Unfair wage	-0.180***	-0.169***	-0.199***	-0.262***
	[0.016]	[0.018]	[0.022]	[0.041]
Net wage/1000	0.054***	0.033***	0.033***	0.032*
	[0.006]	[0.008]	[0.012]	[0.018]
Age	-0.019***	-0.015***	-0.018***	-0.005
	[0.001]	[0.002]	[0.003]	[0.007]
Female	0.013	-0.041*	-0.050*	0.021
	[0.016]	[0.021]	[0.027]	[0.051]
Constant	4.351***	4.334***	4.405***	3.803***
	[0.030]	[0.091]	[0.126]	[0.343]
Further controls	no	yes	yes	yes
Occupational restrictions	no	no	yes	yes
Age restrictions: Age $\geq$ 50	no	no	no	yes
Observations	11,638	9,988	5,892	1,878
R-squared	0.080	0.120	0.132	0.100

Table 5.3: Relation between subjective health status and fairness perceptions (SOEP). OLS estimates with robust standard errors in brackets. The dependent variable measures subjective health status on a five-point scale from "bad" to "very good". ***, **, * indicate significance at the 1-, 5-, and 10-percent level, respectively. "Unfair wage" is a dummy variable equal to one if the respondent answered the question "Do you consider the income that you get at your current job as fair?" with "no" and zero otherwise. Additional controls include marital status (married (baseline category), single, widowed, divorced), whether the respondent lives in East Germany in 2009, labor market status (working in public sector, tenure, full time and part time experience), dummies for educational background (Hauptschule (baseline category), Realschule, Fachoberschulreife, Abitur, other schooling degree, no schooling degree, missing), dummies for firm size (self-employed, below 5, 6-10, 11-20, 21-100 (baseline category), 101-200, 201-2000, more than 2000, missing), occupational status (unskilled blue collar worker, skilled blue collar (baseline category), blue collar craftsman, blue collar foreman, blue collar master, white collar unskilled, white collar skilled, white collar craftsman, white collar master, white collar high qualified, white collar management, civil servant, civil servant intermediate, civil servant high, civil servant executive, other occupation), industry code (agriculture (baseline category), energy, mining, manufacturing, construction, trade, transport, bank/insurance, services, missing). Controls also include measures of personality (Big-5). The sample in column 1 contains all SOEP participants who are in any way active in the labor market in 2009. The sample in column 2 excludes individuals for whom not all controls are available or who just started in the current firm and whose work related information therefore does not refer to the current employer. The sample in column 3 is additionally restricted to dependent full-time employed individuals with positive income. In addition to the restrictions in column 3, the sample in column 4 is restricted to individuals who are at least 50 years old. For more detailed information see Table A5.1 in Appendix A5.

We complement the cross-sectional analysis and exploit the panel structure of the SOEP to develop dynamic panel data models which allow testing for a Granger causal effect of unfair pay on subjective health. Using Arellano-Bover/Blundell-Bond estimators enables us to estimate the model for the years 2011, 2009 and 2007 (for details on data structure, estimation strategy and estimation results see Appendix A5). We estimate and validate models with different lag lengths and robustly reject the null hypothesis that the coefficients of lags of unfair wage perception are zero ($p < 0.05$ in all specifications). This indicates a Granger causal effect of unfair wage perceptions on subjective health. The results are robust for reducing or increasing the lag lengths of subjective health or extending the model by adding lags of net wages.

We now move on to the analysis of specific diseases. Table 5.4 summarizes results from regressions for eight specific diseases listed in the SOEP survey 2009.[15] In addition we constructed a Body Mass Index (BMI) as an additional "objective" health outcome.[16] In Table 5.4 we use the same specifications as in columns (1) to (4) of Table 5.3. Since, with the exception of BMI, outcomes are binary (diagnosed vs. not diagnosed) we use Probit estimates and report average marginal effects. We hypothesized that the unfair wage coefficient should be selectively significant for diseases that are related to stress and impaired cardiac control and especially pronounced for employees older than 50. This is largely what we find: Perceptions of fairness have a highly significant effect on stress-related diseases such as heart disease, high blood pressure, diabetes[17] and high BMI. In contrast, we find only weak or insignificant associations for depression, cancer, asthma, apoplectic stroke or migraine. Comparing columns (3) and (4) in Table 5.4 reveals that the size of the effects concerning heart disease, high blood pressure and diabetes doubles if restricting the sample to employees above age 50. Apparently, and similar to our findings in Table 5.3, ef-

[15]The indication of dementia was also asked for but dementia was excluded from the analysis since less than 0.03% of the working individuals indicated this disease. All regressions are available on request. Note that the data structure of the SOEP does not allow constructing a dynamic panel data model for specific diseases because questions regarding specific diseases were only asked in 2009 and 2011.

[16]BMI is often used as a health indicator, see Currie and Madrian (1999).

[17]The questionnaire asked for diabetes in general, there is no information about different types. Eriksson et al. (2008) suggest that mainly diabetes type II is related to psychological distress.

fects concerning unfair pay and health are driven in particular by older employees. Summarizing, we find selective associations yielding complementary evidence with respect to our findings from the lab.

	Marginal effects of unfair wage			
Disease (Share/mean)	(1)	(2)	(3)	(4)
Heart disease (3.3%)	0.011***	0.013***	0.018***	0.033**
High blood pressure (15.2%)	0.020***	0.019**	0.028***	0.067***
Diabetes (3.2%)	0.008**	0.010***	0.018***	0.033***
Depression (3.9%)	0.008**	0.006	0.007	0.009
Cancer (2.0%)	-0.003	-0.004	0.007*	0.003
Asthma (4.2%)	-0.000	-0.001	0.001	0.016*
Apoplectic stroke (0.5%)	-0.001	-0.001	0.003	0.009
Migraine (5.4%)	0.007	0.006	0.007	0.016*
Body Mass Index (26.0 kg/m^2) (OLS)	0.410***	0.350***	0.305**	0.424*
Further controls	no	yes	yes	yes
Occupational restrictions	no	no	yes	yes
Age restrictions: Age $\geq$ 50	no	no	no	yes

Table 5.4: Relation between specific diseases and unfairness perceptions (SOEP). Regression models (1) to (4) refer to the exact same specifications as in columns (1) to (4) in Table 5.3. We use Probit estimations, reporting average marginal effects, except for Body Mass Index (OLS). Percentages and the BMI mean are related to the full sample in column (1). ***, **, * indicate significance of the "Unfair wage" coefficient at the 1-, 5-, and 10-percent level, respectively.

5.4 Concluding remarks

In this paper we establish a link between the experience of unfair pay and heart rate variability: Higher levels of perceived unfairness go along with lower heart rate variability. Low heart rate variability reflects stress and an impaired balance between the sympathetic and the vagal nervous system, and has been shown to predict coronary heart disease in the long-run. Using a large representative data set (SOEP) we therefore test whether perceptions of unfair pay predict adverse health outcomes in the general population. Our findings suggest that health status is in fact negatively correlated with subjective perceptions of unfair pay. To complement the cross-sectional analysis we exploit the panel dimension of the SOEP, develop dynamic panel data models and provide evidence for a Granger causal effect of unfair pay on health outcomes. Moreover, we find selective associations for specific health outcomes that are predicted if the mechanism operates through the nervous system. Adverse health effects turn out to be most pronounced for full-time employees who are older than 50 years.

Our findings are related to a literature that points out behavioral effects of fairness in labor relations. We show that perceptions of unfair pay not only affect the efficiency of labor relations in reducing work morale (e.g., Fehr et al. (1997)), but also by potentially affecting the health status of the workforce. Our work is also related to research that uses a very different methodological approach: Studies in epidemiology suggest that people who are confined to demanding jobs that fail to compensate efforts by "adequate" rewards are at increased risk of suffering from stress-related disorders (Siegrist, 2005). Other studies suggest that economic inequality in general contributes to adverse health status.[18]

On a more general level our findings provide evidence that the human body registers and systematically processes social and contextual information. This is related, e.g., to findings in Fliessbach et al. (2007) who show that the human brain encodes social comparison. Using fMRI they report that for a given own wage, receiving a wage that is lower than that of another subject is associated with a significantly lower activation in reward-related brain areas, in particular the ventral striatum. In our representative data analysis we show that on top of actual life circumstances and outcomes, such as net wages, mere perceptions of unfair treatment induce adverse physiological responses. Given that health affects labor market outcomes (see, e.g., Currie and Madrian (1999)), this suggests an important potential feedback mechanism: Labor market experience can induce perceptions of unfairness with consequences for health, which in turn affects labor market outcomes. The feedback mechanism between social environment, perceptions and body responses suggests a potential vicious circle and complementary effects. We may thus have to think about some aspects of labor markets differently, with the fairness-health link potentially leading to a vicious circle involving poor pay and poor health. We believe

[18]This was documented in epidemiological investigations using different indicators such as low income (McDonough et al., 1997), income inequality (Kennedy et al., 1996), or perceived unfairness (Bosma et al., 1998; Kivimaeki et al., 2002; Kuper et al., 2002; Lynch et al., 1997). Wilkinson et al. (2011) discuss large-scale effects of inequality.

this question deserved attention in future work.

Bibliography

Abeler, J., Altmann, S., Kube, S. and Wibral, M. (2010), 'Gift exchange and workers' fairness concerns: When equality is unfair', *Journal of the European Economic Association* **8**(6), 1299–1324. [81]

Agell, J. and Lundborg, P. A. (1995), 'Theories of pay and unemployment: Survey evidence from swedish manufacturing firms', *Scandinavian Journal of Economics* **97**(2), 295–307. [81]

Ainsworth, M. S. and Bowlby, J. (1991), 'An ethological approach to personality development.', *American Psychologist* **46**(4), 333 – 341. [50]

Aizer, A. and Currie, J. (2014), 'The intergenerational transmission of inequality: Maternal disadvantage and health at birth', *Science* **344**(6186), 856–861. [80]

Akerlof, G. A. (1982), 'Labor contracts as partial gift exchange', *The Quarterly Journal of Economics* **97**(4), 543–569. [81, 83]

Allport, G. W. and Odbert, H. S. (1936), 'Trait-names: A psycho-lexical study', *Psychological Monographs* **47**(1), i–171. [8, 37]

Almås, I., Cappelen, A. W., Sørensen, E. Ø. and Tungodden, B. (2010), 'Fairness and the development of inequality acceptance', *Science* **328**(5982), 1176–1178. [67]

Almlund, M., Duckworth, A. L., Heckman, J. J. and Kautz, T. D. (2011), 'Personality psychology and economics', *IZA Discussion Paper Series* **5500**. [8, 11, 33, 38, 43, 92]

Altmann, S., Falk, A., Grunewald, A. and Huffman, D. (2014), 'Contractual incompleteness, unemployment, and labour market segmentation', *The Review of Economic Studies* **81**(1), 30–56. [81]

Amthauer, R., Brocke, B., Liepmann, D. and Beauducel, A. (1999), *Intelligenz-Struktur-Test 2000: IST 2000*, Hogrefe & Huber. [40]

Andrews, D. W. K. and Lu, B. (2001), 'Consistent model and moment selection procedures for gmm estimation with application to dynamic panel data models', *Journal of Econometrics* **101**(1), 123–164. [165]

Appelhans, B. M. and Luecken, L. J. (2006), 'Heart rate variability as an index of regulated emotional responding.', *Review of General Psychology* **10**(3), 229. [85]

Arellano, M. and Bond, S. (1991), 'Some tests of specification for panel data: Monte carlo evidence and an application to employment equations', *The Review of Economic Studies* **58**(2), 277–297. [165]

Arrow, K. J. (1972), 'Gifts and exchanges', *Philosophy & Public Affairs* pp. 343–362. [67]

Bandura, A. (1965), 'Influence of models' reinforcement contingencies on the acquisition of imitative responses', *Journal of personality and social psychology* **1**(6), 589–595. [72]

Bandura, A. (1986), *Social foundations of thought and action*, Englewood Cliffs, NJ Prentice Hall. [72]

Bartling, B., Fehr, E., Fischer, B., Kosse, F., Maréchal, M., Pfeiffer, F., Schunk, D., Schupp, J., Spieß, C. K. and Wagner, G. G. (2010), 'Determinanten kindlicher Geduld – Ergebnisse einer Experimentalstudie im Haushaltskontext', *Schmollers Jahrbuch* **130**, 297–323. [37, 53, 59, 60]

Bauer, M., Chytilová, J. and Pertold-Gebicka, B. (2014), 'Parental background and other-regarding preferences in children', *Experimental Economics* **17**(1), 24–46. [72]

Becker, A., Deckers, T., Dohmen, T., Falk, A. and Kosse, F. (2012), 'The relationship between economic preferences and psychological personality measures', *Annual Review of Economics* **4**(1), 453–478. [2, 37, 38, 39, 49, 54, 63, 64, 92, 150]

Becker, G. S. (2009), *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, University of Chicago Press. [35]

Ben-Ner, A. and Kramer, A. (2011), 'Personality and altruism in the dictator game: Relationship to giving to kin, collaborators, competitors, and neutrals', *Personality and Individual Differences* **51**(3), 216–221. [12]

Berg, J., Dickhaut, J. and McCabe, K. (1995), 'Trust, reciprocity, and social history', *Games and Economic Behavior* **10**(1), 122–142. [17]

Bettinger, E. P., Long, B. T., Oreopoulos, P. and Sanbonmatsu, L. (2012), 'The role of application assistance and information in college decisions: Results from the h&r block fafsa experiment', *The Quarterly Journal of Economics* **127**(3), 1205–1242. [38]

Bettinger, E. and Slonim, R. (2007), 'Patience among children', *Journal of Public Economics* **91**(1), 343–363. [37]

Bewley, T. F. (1999), *Why Wages Don't Fall During a Recession*, Harvard University Press. [81]

Bewley, T. F. (2005), *Moral sentiments and material interests: The foundations of cooperation in economic life*, The MIT Press, Cambridge, Massachusetts, chapter Fairness, Reciprocity, and Wage Rigidity, pp. 303–338. [81]

Bibby, P. A. and Ferguson, E. (2011), 'The ability to process emotional information predicts loss aversion', *Personality and Individual Differences* **51**(3), 263–266. [12]

Bisin, A. and Verdier, T. (2001), 'The economics of cultural transmission and the dynamics of preferences', *Journal of Economic Theory* **97**(2), 298–319. [72]

Blomeyer, D., Coneus, K., Laucht, M. and Pfeiffer, F. (2009), 'Initial risk matrix, home resources, ability development, and children's achievement', *Journal of the European Economic Association* **7**(2-3), 638–648. [51, 136]

Borghans, L., Duckworth, A. L., Heckman, J. J. and Ter Weel, B. (2008), 'The economics and psychology of personality traits', *Journal of Human Resources* **43**(4), 972–1059. [8, 37, 118]

Borghans, L., Golsteyn, B. H., Heckman, J. J. and Humphries, J. E. (2011), 'Identification problems in personality psychology', *Personality and Individual Differences* **51**(3), 315–320. [40]

Borghans, L., Heckman, J. J., Golsteyn, B. H. and Meijers, H. (2009), 'Gender differences in risk aversion and ambiguity aversion', *Journal of the European Economic Association* **7**(2-3), 649–658. [12]

Bosma, H., Peter, R., Siegrist, J. and Marmot, M. (1998), 'Two alternative job stress models and the risk of coronary heart disease', *American Journal of Public Health* **88**(1), 68–74. [96]

Bradley, R. H. and Caldwell, B. M. (1981), 'The home inventory: A validation of the preschool scale or black children', *Child Development* pp. 708–710. [51, 136]

Brosschot, J. F. and Thayer, J. F. (2003), 'Heart rate response is longer after negative emotions than after positive emotions', *International Journal of Psychophysiology* **50**(3), 181–187. [84]

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E. and Pan, Y. (2014), 'Early childhood investments substantially boost adult health', *Science* **343**(6178), 1478–1485. [80]

Card, D. (1995), *Myth and Measurement: The New Economics of the Minimum Wage*, Princeton University Press. [81]

Card, D. (2001), 'Estimating the return to schooling: Progress on some persistent econometric problems', *Econometrica* **69**(5), 1127–1160. [35]

Carneiro, P. (2003), Heterogeneity in the Returns to Schooling: Implications for Policy Evaluation, PhD thesis, University of Chicago, Department of Economics August 2003. [35]

Cattell, R. B. (1971), *Abilities: Their Structure, Growth, and Action*, Houghton Mifflin. [53]

Cattell, R. B. (1987), *Intelligence: Its Structure, Growth and Action: Its Structure, Growth and Action*, Elsevier. [36]

Charness, G. and Kuhn, P. (2011), *Handbook of Labor Economics*, Vol. 4 A, Elsevier, chapter Lab labor: What can Labor Economists Learn from the Lab?, pp. 229–330. [81]

Chen, A. and Rogan, W. J. (2004), 'Breastfeeding and the risk of postneonatal death in the united states', *Pediatrics* **113**(5), 435–439. [50]

Cobb-Clark, D. A. and Tan, M. (2011), 'Noncognitive skills, occupational attainment, and relative wages', *Labour Economics* **18**(1), 1–13. [36]

Cohen, J. (1988), *Statistical power analysis for the behavioral sciences.*, 2 edn, Hillsdale, New Jersey: Lawrence Erlbaum Associates. [25, 26]

Cohen, J. (1992), 'A power primer.', *Psychological Bulletin* **112**(1), 155–159. [22]

Conti, G. and Heckman, J. J. (2010), 'Understanding the early origins of the education–health gradient a framework that can also be applied to analyze gene–environment interactions', *Perspectives on Psychological Science* **5**(5), 585–605. [35, 36, 92]

Costa, P. T. and McCrae, R. R. (1989), *Neo PI/FFI Manual Supplement*, Psychological Assessment Resources: Odessa, FL. [9, 18]

Costa, P. T. and McCrae, R. R. (1992), *Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI)*, Psychological Assessment Resources: Odessa, FL. [37]

Cunha, F. and Heckman, J. (2007), 'The technology of skill formation', *American Economic Review* **97**(2), 31–47. [35, 50, 56, 67]

Cunha, F. and Heckman, J. J. (2008), 'Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation', *Journal of Human Resources* **43**(4), 738–782. [35, 67]

Currie, J. and Almond, D. (2011), *Handbook of Labor Economics*, Vol. 4 B, Elsevier, chapter Human Capital Development Before Age Five, pp. 1315–1486. [50]

Currie, J. and Madrian, B. C. (1999), *Handbook of Labor Economics*, Vol. 3, Elsevier, chapter Health, Health Insurance and the Labor Market, pp. 3309–3416. [94, 96]

Currie, J. and Moretti, E. (2003), 'Mother's education and the intergenerational transmission of human capital: Evidence from college openings', *The Quarterly Journal of Economics* **118**(4), 1495–1532. [53]

Daly, M., Harmon, C. P. and Delaney, L. (2009), 'Psychological and biological foundations of time preference', *Journal of the European Economic Association* **7**(2-3), 659–669. [11]

Dekker, J. M., Crow, R. S., Folsom, A. R., Hannan, P. J., Liao, D., Swenne, C. A. and Schouten, E. G. (2000), 'Low heart rate variability in a 2-minute rhythm strip predicts risk of coronary heart disease and mortality from several causes the aric study', *Circulation* **102**(11), 1239–1244. [81, 85]

Doepke, M. and Zilibotti, F. (2008), 'Occupational choice and the spirit of capitalism', *The Quarterly Journal of Economics* **123**(2), 747–793. [57]

Doepke, M. and Zilibotti, F. (2012), 'Parenting with style: Altruism and paternalism in intergenerational preference transmission', *IZA Discussion Paper Series* **7108**. [57]

Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2008), 'Representative trust and reciprocity: Prevalence and determinants', *Economic Inquiry* **46**(1), 84–90. [12]

Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2009), 'Homo reciprocans: Survey evidence on behavioural outcomes', *The Economic Journal* **119**(536), 592–612. [21]

Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2010), 'Are risk aversion and impatience related to cognitive ability?', *The American Economic Review* **100**(3), 1238–1260. [10, 11, 12, 15, 19, 53, 54, 58, 61, 62]

Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2012), 'The intergenerational transmission of risk and trust attitudes', *The Review of Economic Studies* **79**(2), 645–677. [72]

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. and Wagner, G. G. (2011), 'Individual risk attitudes: Measurement, determinants, and behavioral consequences', *Journal of the European Economic Association* **9**(3), 522–550. [21, 33, 39, 49, 64]

Dunifon, R., Duncan, G. J. and Brooks-Gunn, J. (2001), 'As ye sweep, so shall ye reap', *American Economic Review* pp. 150–154. [37]

Eckel, C. C. and Grossman, P. J. (1996), 'Altruism in anonymous dictator games', *Games and Economic Behavior* **16**(2), 181–191. [7, 18]

Eckel, C. C. and Grossman, P. J. (2002), 'Sex differences and statistical stereotyping in attitudes toward financial risk', *Evolution and Human Behavior* **23**(4), 281–295. [12]

Eriksson, A.-K., Ekbom, A., Granath, F., Hilding, A., Efendic, S. and Östenson, C.-G. (2008), 'Psychological distress and risk of pre-diabetes and type 2 diabetes in a prospective study of swedish middle-aged men and women', *Diabetic Medicine* **25**(7), 834–842. [94]

Faircloth, C. (2014), *Parenting Culture Studies*, Palgrave MacMillan, chapter Intensive Parenting and the Expansion of Parenting, pp. 25–50. [63]

Falk, A., Becker, A., Dohmen, T., Huffman, D. and Sunde, U. (2011), 'An experimentally validated preference module', *Unpublished manuscript* . [14, 150]

Falk, A., Fehr, E. and Fischbacher, U. (2005), 'Driving forces behind informal sanctions', *Econometrica* **73**(6), 2017–2030. [18]

Falk, A., Fehr, E. and Zehnder, C. (2006), 'Fairness perceptions and reservation wages-the behavioral effects of minimum wage laws', *The Quarterly Journal of Economics* **121**(4), 1347–1381. [81]

Falk, A. and Fischbacher, U. (2006), 'A theory of reciprocity', *Games and Economic Behavior* **54**(2), 293–315. [7, 86]

Falk, A. and Heckman, J. J. (2009), 'Lab experiments are a major source of knowledge in the social sciences', *Science* **326**(5952), 535–538. [90]

Farkas, G. (2003), 'Cognitive skills and noncognitive traits and behaviors in stratification processes', *Annual Review of Sociology* pp. 541–562. [36]

Fehr, E. (2009), 'On the economics and biology of trust', *Journal of the European Economic Association* **7**(2-3), 235–266. [7]

Fehr, E., Bernhard, H. and Rockenbach, B. (2008), 'Egalitarianism in young children', *Nature* **454**(7208), 1079–1083. [53, 59, 67, 70]

Fehr, E., Fischbacher, U., von Rosenbladt, B., Schupp, J. and Wagner, G. G. (2002), 'A nation-wide laboratory', *Schmollers Jahrbuch* **122**(519), 542. [71, 148, 150]

Fehr, E. and Gächter, S. (2000), 'Fairness and retaliation: The economics of reciprocity', *The Journal of Economic Perspectives* pp. 159–181. [81]

Fehr, E. and Gächter, S. (2002), 'Altruistic punishment in humans', *Nature* **415**(6868), 137–140. [67]

Fehr, E., Gächter, S. and Kirchsteiger, G. (1997), 'Reciprocity as a contract enforcement device: Experimental evidence', *Econometrica* pp. 833–860. [67, 81, 96]

Fehr, E., Glätzle-Rützler, D. and Sutter, M. (2013), 'The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence', *European Economic Review* **64**, 369–383. [67]

Fehr, E., Kirchsteiger, G. and Riedl, A. (1993), 'Does fairness prevent market clearing? an experimental investigation', *The Quarterly Journal of Economics* **108**(2), 437–459. [81]

Fehr, E. and Schmidt, K. M. (1999), 'A theory of fairness, competition, and cooperation', *The Quarterly Journal of Economics* **114**(3), 817–868. [86]

Fischbacher, U. (2007), 'z-tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics* **10**(2), 171–178. [15, 84]

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E. and Falk, A. (2007), 'Social comparison affects reward-related brain activity in the human ventral striatum', *Science* **318**(5854), 1305–1308. [96]

Frederick, S., Loewenstein, G. and O'Donoghue, T. (2002), 'Time discounting and time preference: A critical review', *Journal of Economic Literature* **40**(2), 351–401. [7, 37]

Gauthier, A. H., Smeeding, T. M. and Furstenberg, F. F. (2004), 'Are parents investing less time in children? trends in selected industrialized countries', *Population and Development Review* **30**(4), 647–672. [63]

Gerlitz, J.-Y. and Schupp, J. (2005), 'Zur Erhebung der Big-Five-basierten Persoenlichkeitsmerkmale im SOEP', *DIW Research Notes* **4**. [9, 19, 22, 39]

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S. M. and Grantham-McGregor, S. (2014), 'Labor market returns to an early childhood stimulation intervention in Jamaica', *Science* **344**(6187), 998–1001. [80]

Gianaros, P. J., Salomon, K., Zhou, F., Owens, J. F., Edmundowicz, D., Kuller, L. H. and Matthews, K. A. (2005), 'A greater reduction in high-frequency heart rate variability to a psychological stressor is associated with subclinical coronary and aortic calcification in postmenopausal women', *Psychosomatic Medicine* **67**(4), 553–560. [81, 85]

Gollier, C. (2004), *The Economics of Risk and Time*, MIT press. [7]

Goodman, R. (1997), 'The strengths and difficulties questionnaire: a research note', *Journal of child psychology and psychiatry* **38**(5), 581–586. [71]

Granger, C. W. (1969), 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica* pp. 424–438. [164]

Greiner, B. (2004), 'An online recruitment system for economic experiments', *Forschung und wissenschaftliches Rechnen* **63**, 79–93. [15]

Guiso, L., Sapienza, P. and Zingales, L. (2009), 'Cultural biases in economic exchange?', *The Quarterly Journal of Economics* **124**(3), 1095–1131. [67]

Hamilton, J. D. (1994), *Time Series Analysis*, Vol. 2, Princeton university press Princeton. [164]

Haushofer, J. and Fehr, E. (2014), 'On the psychology of poverty', *Science* **344**(6186), 862–867. [80]

Hays, S. (1996), *The Cultural Contradictions of Motherhood*, Yale University Press. [63]

Heckman, J. J. (1979), 'Sample selection bias as a specification error', *Econometrica: Journal of the Econometric Society* pp. 153–161. [35]

Heckman, J. J. (2006), 'Skill formation and the economics of investing in disadvantaged children', *Science* **312**(5782), 1900–1902. [50, 56, 67]

Heckman, J. J., Humphries, J. E., Urzua, S. and Veramendi, G. (2013), 'The effects of educational choices on labor market, health, and social outcomes', *Unpublished manuscript, University of Chicago, Department of Economics* . [36, 38]

Heckman, J. J. and Kautz, T. (2014), Fostering and measuring skills: Interventions that improve character and cognition, *in* J. J. Heckman, J. E. Humphries and T. Kautz, eds, 'The Myth of Achievement Tests', University of Chicago Press, pp. 341–430. [69, 146]

Heckman, J. J. and Rubinstein, Y. (2001), 'The importance of noncognitive skills: Lessons from the ged testing program', *American Economic Review* pp. 145–149. [35]

Heckman, J. J., Stixrud, J. and Urzua, S. (2006), 'The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior', *Journal of Labor Economics* **24**(3), pp. 411–482. [35, 36, 40, 49]

Heimerdinger, T. (2009), *Bilder. Bücher. Bytes*, Waxmann: Münster, chapter Brust oder Flasche?–Säuglingsernährung und die Rolle von Beratungsmedien, pp. 100–110. [57, 63]

Heinrichs, M., Neumann, I. and Ehlert, U. (2002), 'Lactation and stress: Protective effects of breast-feeding in humans', *Stress: The International Journal on the Biology of Stress* **5**(3), 195–203. [50]

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E. and Gintis, H. (2004), *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, Oxford University Press. [80]

Herrnstein, R. J. and Murray, C. (2010), *Bell Curve: Intelligence and Class Structure in American Life*, Simon and Schuster. [35]

Holtz-Eakin, D., Newey, W. and Rosen, H. S. (1988), 'Estimating vector autoregressions with panel data', *Econometrica* pp. 1371–1395. [164]

Ida, T. and Goto, R. (2009), 'Simultaneous measurement of time and risk preferences: Stated preference discrete choice modeling analysis depending on smoking behavior', *International Economic Review* **50**(4), 1169–1182. [56]

Imbens, G. W. and Wooldridge, J. M. (2009), 'Recent developments in the econometrics of program evaluation', *Journal of Economic Literature* **47**(1), 5–86. [80]

Jacob, B. A. (2002), 'Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education', *Economics of Education Review* **21**(6), 589–598. [35, 36]

Jayachandran, S. and Kuziemko, I. (2011), ‘Why do mothers breastfeed girls less than boys? evidence and implications for child health in india’, *The Quarterly Journal of Economics* **126**(3), 1485–1538. [50, 53]

Kaestner, R. and Callison, K. (2011), ‘Adolescent cognitive and noncognitive correlates of adult health’, *Journal of Human Capital* **5**(1), 29–69. [37]

Katz, L. and Krueger, A. (1992), ‘The effect of the minimum wage on the fast-food industry’, *Industrial & Labor Relations Review* **46**(1), 6–21. [81]

Kennedy, B. P., Kawachi, I. and Prothrow-Stith, D. (1996), ‘Income distribution and mortality: Cross sectional ecological study of the robin hood index in the united states’, *British Medical Journal* **312**(7037), 1004–1007. [96]

KiGGS (2008), ‘German Health Survey for Children and Adolescents, Robert Koch Institute, Berlin’. [56, 140]

Kivimaeki, M., Leino-Arjas, P., Luukkonen, R., Riihimaeki, H., Vahtera, J. and Kirjonen, J. (2002), ‘Work stress and risk of cardiovascular mortality: Prospective cohort study of industrial employees’, *British Medical Journal* **325**(7369), 857. [96]

Knack, S. and Keefer, P. (1997), ‘Does social capital have an economic payoff? a cross-country investigation’, *The Quarterly journal of economics* pp. 1251–1288. [67]

Koeszegi, B. (2006), ‘Ego utility, overconfidence, and task choice’, *Journal of the European Economic Association* **4**(4), 673–707. [34]

Kosse, F. and Pfeiffer, F. (2012), ‘Impatience among preschool children and their mothers’, *Economics Letters* **115**(3), 493–495. [37, 72]

Kube, S., Maréchal, M. A. and Puppe, C. (2012), ‘The currency of reciprocity: Gift exchange in the workplace’, *The American Economic Review* **102**(4), 1644–1662. [81]

Kuper, H., Singh-Manoux, A., Siegrist, J. and Marmot, M. (2002), 'When reciprocity fails: Effort–reward imbalance in relation to coronary heart disease and health functioning within the whitehall ii study', *Occupational and Environmental Medicine* **59**(11), 777–784. [96]

La Porta, R., Lopez-De-Silanes, F., Shleifer, A. and Vishny, R. W. (1997), 'Legal determinants of external finance', *Journal of Finance* pp. 1131–1150. [67]

Lang, F. R., Hahne, D., Gymbel, S., Schröpper, S. and Lutsch, K. (2005), 'Erfassung des kognitiven Leistungspotenzials und der "Big Five" mit Computer-Assisted-Personal-Interviewing (CAPI): zur Reliabilität und Validität zweier ultrakurzer Tests und des BFI-S', *DIW Research Notes* **9**. [60, 61]

Lang, F. R., Weiss, D., Stocker, A. and von Rosenbladt, B. (2007), 'Assessing cognitive capacities in computer-assisted survey research: Two ultra-short tests of intellectual ability in the german socio-economic panel (soep)', *Schmollers Jahrbuch* **127**(1), 183–192. [61]

Lanigan, J., Singhal, A. et al. (2009), 'Early nutrition and long-term health: A practical approach', *Proceedings of the Nutrition Society* **68**(4), 422. [50]

Li, R., Scanlon, K. S. and Serdula, M. K. (2005), 'The validity and reliability of maternal recall of breastfeeding practice', *Nutrition Reviews* **63**(4), 103–110. [52]

Liu, J. L. Y., Maniadakis, N., Gray, A. and Rayner, M. (2002), 'The economic burden of coronary heart disease in the UK', *Heart* **88**(6), 597–603. [83]

Lleras, C. (2008), 'Do skills and behaviors in high school matter? the contribution of noncognitive factors in explaining differences in educational attainment and earnings', *Social Science Research* **37**(3), 888–902. [36, 37]

Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T., Murray, C., Haisma, H., Coward, W., Visser, G., Vonk, R., Wells, J. et al. (2006), 'Global burden of disease and risk factors', *Journal of Nutrition* **136**(11), 2945–2951. [83]

Lynch, J. W., Kaplan, G. A. and Shema, S. J. (1997), 'Cumulative impact of sustained economic hardship on physical, cognitive, psychological, and social functioning', *New England Journal of Medicine* **337**(26), 1889–1895. [96]

Manski, C. F. (2002), 'Identification of decision rules in experiments on simple games of proposal and response', *European Economic Review* **46**(4–5), 880 – 891. [13]

McDonough, P., Duncan, G. J., Williams, D. and House, J. (1997), 'Income dynamics and adult mortality in the united states, 1972 through 1989', *American Journal of Public Health* **87**(9), 1476–1483. [96]

McKenzie, D. (2012), 'Beyond baseline and follow-up: The case for more t in experiments', *Journal of Development Economics* **99**(2), 210–221. [80]

Meyer-Lindenberg, A., Domes, G., Kirsch, P. and Heinrichs, M. (2011), 'Oxytocin and vasopressin in the human brain: Social neuropeptides for translational medicine', *Nature Reviews Neuroscience* **12**(9), 524–538. [50]

Michaud, P.-C. and Van Soest, A. (2008), 'Health and wealth of elderly couples: Causality tests using dynamic panel data models', *Journal of Health Economics* **27**(5), 1312–1325. [164]

Mischel, W., Shoda, Y. and Rodriguez, M. I. (1989), 'Delay of gratification in children', *Science* **244**(4907), 933–938. [53, 59]

Neal, D. A. and Johnson, W. R. (1996), 'The role of premarket factors in black-white wage differences', *The Journal of Political Economy* **104**(5), 869–895. [35]

Nielsen, H. S., Sørensen, T. and Taber, C. (2010), 'Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform', *American Economic Journal: Economic Policy* **2**(2), 185–215. [38]

Niskanen, J.-P., Tarvainen, M. P., Ranta-Aho, P. O. and Karjalainen, P. A. (2004), 'Software for advanced hrv analysis', *Computer Methods and Programs in Biomedicine* **76**(1), 73–81. [85]

Ostrom, E., Dietz, T., Dolsak, N., Stern, P. C., Stonich, S. and Weber, E. U. (2002), *The drama of the commons*, National Academies Press. [67]

Penttilae, J., Helminen, A., Jartti, T., Kuusela, T., Huikuri, H. V., Tulppo, M. P., Coffeng, R. and Scheinin, H. (2001), 'Time domain, geometrical and frequency domain analysis of cardiac vagal outflow: Effects of various respiratory patterns', *Clinical Physiology* **21**(3), 365–376. [85]

Piketty, T. and Saez, E. (2014), 'Inequality in the long run', *Science* **344**(6186), 838–843. [80]

Richerson, P. J. and Boyd, R. (2008), *Not by genes alone: How culture transformed human evolution*, University of Chicago Press. [80]

Roberts, B. W. (2006), 'Personality development and organizational behavior', *Research in Organizational Behavior* **27**, 1–40. [8]

Roberts, B. W. (2009), 'Back to the future: Personality and assessment and personality development', *Journal of Research in Personality* **43**(2), 137–145. []

Rodríguez-Planas, N. (2012), 'Mentoring, educational services, and incentives to learn: What do we know about them?', *Evaluation and program planning* **35**(4), 481–490. [69, 146]

Roger, V. L., Go, A. S., Lloyd-Jones, D. M., Benjamin, E. J., Berry, J. D., Borden, W. B., Bravata, D. M., Dai, S., Ford, E. S., Fox, C. S. et al. (2012), 'Heart disease and stroke statistics-2012 update a report from the american heart association', *Circulation* **125**(1), 2–220. [83, 91, 165]

Rotter, J. B. (1966), 'Generalized expectancies for internal versus external control of reinforcement', *Psychological monographs: General and Applied* **80**(1), 1. [8, 9, 19, 22, 40]

Rustichini, A., DeYoung, C. G., Anderson, J. and Burks, S. V. (2012), 'Toward the integration of personality theory and decision theory in the explanation of economic and health behavior', *Discussion Paper series* **6750**. [11, 12, 36, 38]

Samuelson, P. A. (1937), 'A note on measurement of utility', *The Review of Economic Studies* **4**(2), 155–161. [7]

Sayer, L. C., Bianchi, S. M. and Robinson, J. P. (2004), 'Are parents investing less in children? trends in mothers' and fathers' time with children', *American Journal of Sociology* **110**(1), 1–43. [63]

Schupp, J. and Wagner, G. G. (2002), 'Maintenance of and innovation in long-term panel studies: The case of the german socio-economic panel (GSOEP)', *Allgemeines Statistisches Archiv* **86**(2), 163–175. [20, 82, 90]

Selten, R. (1967), *Die Strategiemethode zur Erforschung des eingeschraenkt rationalen Verhaltens im Rahmen eines Oligopolexperiments*, J.C.B. Mohr (Paul Siebeck). [17]

Siegrist, J. (2005), 'Social reciprocity and health: New scientific evidence and policy implications', *Psychoneuroendocrinology* **30**(10), 1033–1038. [96]

Slovic, P. (1966), 'Risk-taking in children: Age and sex differences', *Child Development* pp. 169–176. [53, 59]

SOEP (2012), 'German Socio-economic Panel, Version 28, Deutsches Institut für Wirtschaftsforschung, Berlin'. [50, 57, 63, 134, 135, 136, 141, 142]

Solga, H., Stern, E., von Rosenbladt, B., Schupp, J. and Wagner, G. G. (2005), 'The measurement and importance of general reasoning potentials in schools and labor markets', *Pre-Test Report. Research Notes* (10). [40]

Sroufe, L. A. (1988), *Clinical implications of attachment*, Lawrence Erlbaum, Hillsdale, NJ, chapter The Role of Infant-Caregiver Attachment in Development, pp. 18–38. [50]

Steptoe, A. and Marmot, M. (2002), 'The role of psychobiological pathways in socio-economic inequalities in cardiovascular disease risk', *European Heart Journal* **23**(1), 13–25. [81, 85]

Sutter, M. and Kocher, M. G. (2007), 'Trust and trustworthiness across different age groups', *Games and Economic Behavior* **59**(2), 364–382. [67]

Sutter, M., Kocher, M. G., Glatzle-Rutzler, D. and Trautmann, S. T. (2013), 'Impatience and uncertainty: Experimental decisions predict adolescents' field behavior', *The American Economic Review* **103**(1), 510–531. [49, 56]

Task Force (1996), 'Heart rate variability: Standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology', *Circulation* **93**(5), 1043–1065. [85]

Tewes, U. (1994), *Hamburg Wechsler Test für Erwachsene (HAWIE-R)*, Bern: Huber. [60]

Uauy, R. and De Andraca, I. (1995), 'Human milk and breast feeding for optimal mental development', *Journal of Nutrition* **125**(8), 2278S. [50]

Van der Klaauw, W. (2002), 'Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach', *International Economic Review* **43**(4), 1249–1287. [38]

Vermunt, R. and Steensma, H. (2003), 'Physiological relaxation: Stress reduction through fair treatment', *Social Justice Research* **16**(2), 135–149. [85]

Vischer, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U. and Wagner, G. G. (2013), 'Validating an ultra-short survey measure of patience', *Economics Letters* **120**(2), 142–145. [21, 40, 64]

von Borell, E., Langbein, J., Després, G., Hansen, S., Leterrier, C., Marchant-Forde, J., Marchant-Forde, R., Minero, M., Mohr, E., Prunier, A. et al. (2007), 'Heart rate variability as a measure of autonomic regulation of cardiac activity for assessing stress and welfare in farm animals-a review', *Physiology & Behavior* **92**(3), 293–316. [85]

Waddell, G. R. (2006), 'Labor-market consequences of poor attitudwe and low self-esteem in youth', *Economic Inquiry* **44**(1), 69–97. [37]

Wagner, G. G., Frick, J. R. and Schupp, J. (2007), 'The german socio-economic panel study (soep)–scope, evolution and enhancements', *Schmollers Jahrbuch:*

Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts-und Sozialwissenschaften **127**(1), 139–169. [20, 36, 38, 50, 57, 58, 63]

Weiss, R. (2006), *Grundintelligenztest (CFT 20-R)*, Goettingen: Hogrefe. [60]

Wilkinson, R. G., Pickett, K. and Chafer, C. (2011), *The Spirit Level*, Tantor Media, Incorporated. [96]

Willis, R. J. and Rosen, S. (1979), 'Education and self-selection', *Journal of Political Economy* **87**(5 pt 2). [35]

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, MIT press. [52]

Zak, P. J. and Knack, S. (2001), 'Trust and growth', *The economic journal* **111**(470), 295–321. [67]

Zuckerman, M. (1994), *Behavioral Expressions and Biosocial Bases of Sensation Seeking*, Cambridge University Press. [12]

Appendices

A1 Appendix to Chapter 1

Table A1.1: Definitions of the Big Five domains

Big Five Domain	APA Dictionary Definition
Openness	Refers to individual differences in the tendency to be open to new aesthetic, cultural, or intellectual experiences.
Conscientiousness	The tendency to be organized, responsible, and hardworking; one end of a dimension of individual differences: conscientiousness vs. lack of direction.
Extraversion	An orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience. Extroverts are relatively more outgoing, gregarious, sociable, and openly expressive.
Agreeableness	The tendency to act in a cooperative, unselfish manner; one end of a dimension of individual differences: agreeableness vs. disagreeableness.
Neuroticism	Characterized by a chronic level of emotional instability and proneness to psychological distress.

This table is in parts reproduced from Borghans et al. (2008).

Table A1.2: Spearman correlation structure experimental data set

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism	LoC
Time	0.0388	0.0162	−0.0114	0.1077**	−0.0684	0.1063*
Risk	0.0027	−0.0486	0.0786*	0.0206	−0.0995**	0.0485
Pos. Reciprocity	0.1606***	0.0078	0.0177	0.2029***	0.0152	0.0441
Neg. Reciprocity	−0.0967*	−0.0221	0.0462	−0.083*	−0.0165	−0.1376**
Trust	0.1354***	−0.1198***	0.002	0.1696***	−0.002	−0.0648
Altruism	0.0969*	−0.0804	0.0034	0.2000***	0.0879*	0.0418

*, **, and *** indicate significance at the 10%, 5%, and 1% level. Correlations between economic preferences and the Big Five were calculated using between 394 and 477 observations. Correlations between economic preferences and Locus of Control were calculated using between 254 and 315 observations. All measures are standardized.

Table A1.3: Spearman correlation structure representative experimental data

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Time	−0.0199	−0.0737	−0.0764*	−0.0829*	−0.0598
Risk	0.1315*	−0.0744	0.0661	−0.0854*	−0.0261

*, **, and *** indicate significance at the 10%, 5%, and 1% level. All measures are standardized.

Table A1.4: Spearman correlation structure SOEP

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism	LoC
Time	0.0233	0.1192	−0.0342	0.3099	−0.0643	0.0709
Risk	0.2632	−0.0500	0.2452	−0.1496	−0.1049	0.1426
Pos. Reciprocity	0.1835	0.2622	0.1547	0.1947	0.0808	0.1041
Neg. Reciprocity	−0.0616	−0.1767	−0.0426	−0.3853	0.0572	−0.2257
Trust	0.1224	−0.0693	0.0523	0.0788	−0.1889	0.2012
Altruism	0.1693	0.1501	0.1602	0.2416	0.0860	0.0843

All correlations are significant at the 1% level and are calculated using 14,243 observations. All measures are standardized.

Table A1.5: Outcome regressions: Representative experimental data

	(1)	(2)	(3)	(4)	(5)
Life Outcomes	Subj. Health	Life Satisf.	Gross Wage	Unemployed	Years of Educ.
Openness	0.043***	0.123***	0.989***	-0.018***	0.667***
	(0.009)	(0.017)	(0.162)	(0.004)	(0.027)
Conscientiousn.	0.038***	0.106***	0.565***	-0.014***	-0.182***
	(0.009)	(0.017)	(0.161)	(0.004)	(0.026)
Extraversion	0.026***	0.134***	-1.201***	0.006*	-0.309***
	(0.009)	(0.017)	(0.154)	(0.004)	(0.026)
Agreeableness	0.033***	0.139***	-1.288***	0.023***	-0.146***
	(0.010)	(0.018)	(0.165)	(0.004)	(0.028)
Neuroticism	-0.140***	-0.186***	-1.009***	0.018***	-0.272***
	(0.009)	(0.016)	(0.158)	(0.004)	(0.026)
LoC	0.105***	0.307***	1.899***	-0.043***	0.421***
	(0.008)	(0.015)	(0.145)	(0.003)	(0.024)
Patience	0.024***	0.129***	-0.343**	0.001	-0.151***
	(0.008)	(0.015)	(0.136)	(0.003)	(0.023)
Risk	0.131***	0.076***	0.415**	0.003	0.210***
	(0.009)	(0.017)	(0.166)	(0.004)	(0.027)
Pos. Recip.	-0.035***	0.006	0.388***	-0.002	0.005
	(0.008)	(0.015)	(0.140)	(0.003)	(0.023)
Neg. Recip.	0.064***	0.039**	-0.329**	0.006*	-0.137***
	(0.008)	(0.015)	(0.147)	(0.003)	(0.024)
Trust	0.122***	0.308***	1.763***	-0.035***	0.587***
	(0.009)	(0.015)	(0.145)	(0.003)	(0.024)
Altruism	0.070***	0.072***	-0.780***	0.005	0.084***
	(0.009)	(0.016)	(0.152)	(0.003)	(0.025)
Constant	3.300***	6.852***	16.100***	0.099***	12.346***
	(0.007)	(0.014)	(0.131)	(0.003)	(0.021)
Observations	14,218	14,214	7,199	9,095	13,768
Adj. R-squared	0.108	0.159	0.0919	0.0547	0.174

*, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. All measures are standardized.

Figure A1.1: Kernel-weighted local linear polynomial regressions using experimental data

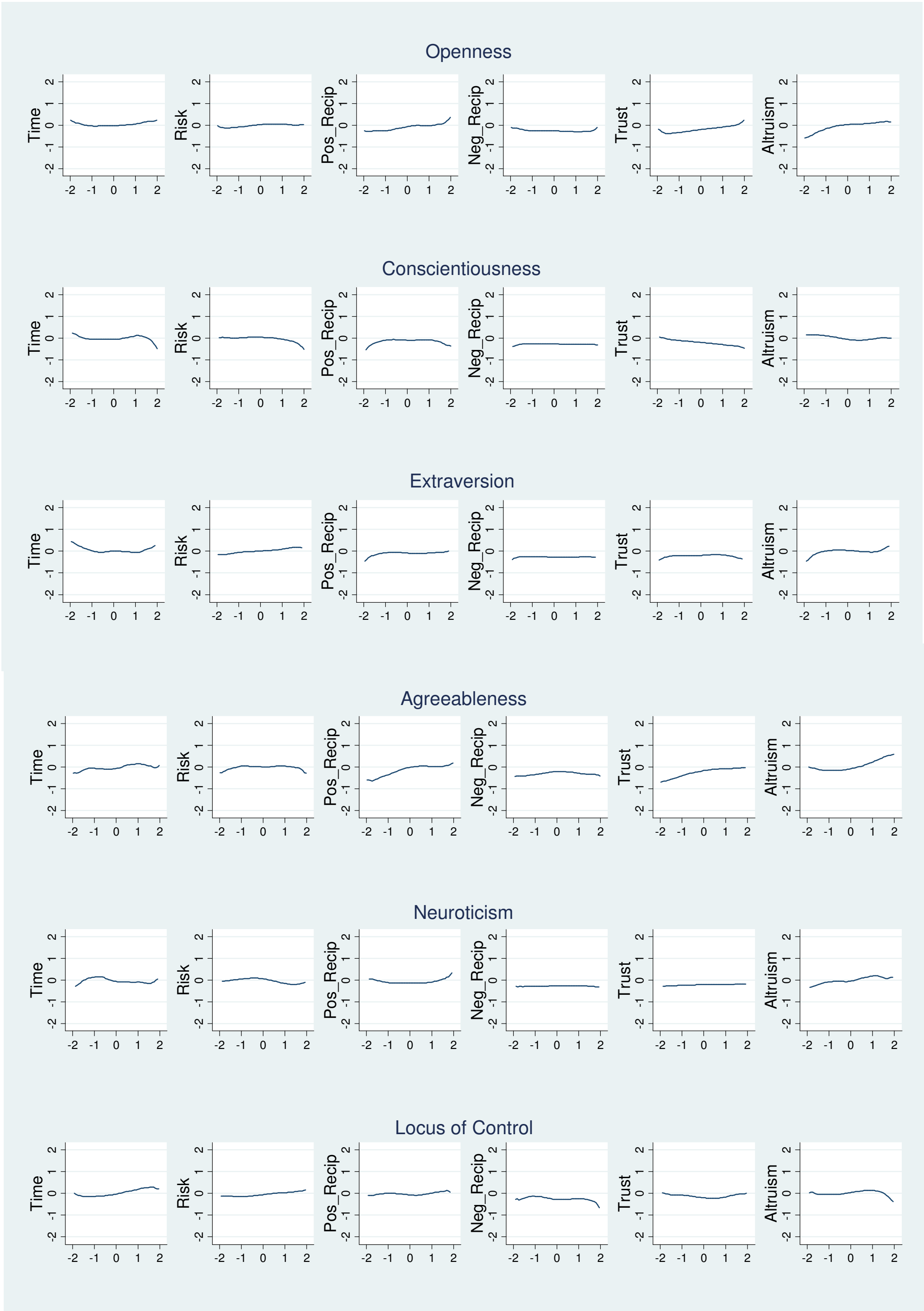


Figure A1.2: Kernel-weighted local linear polynomial regressions using SOEP data

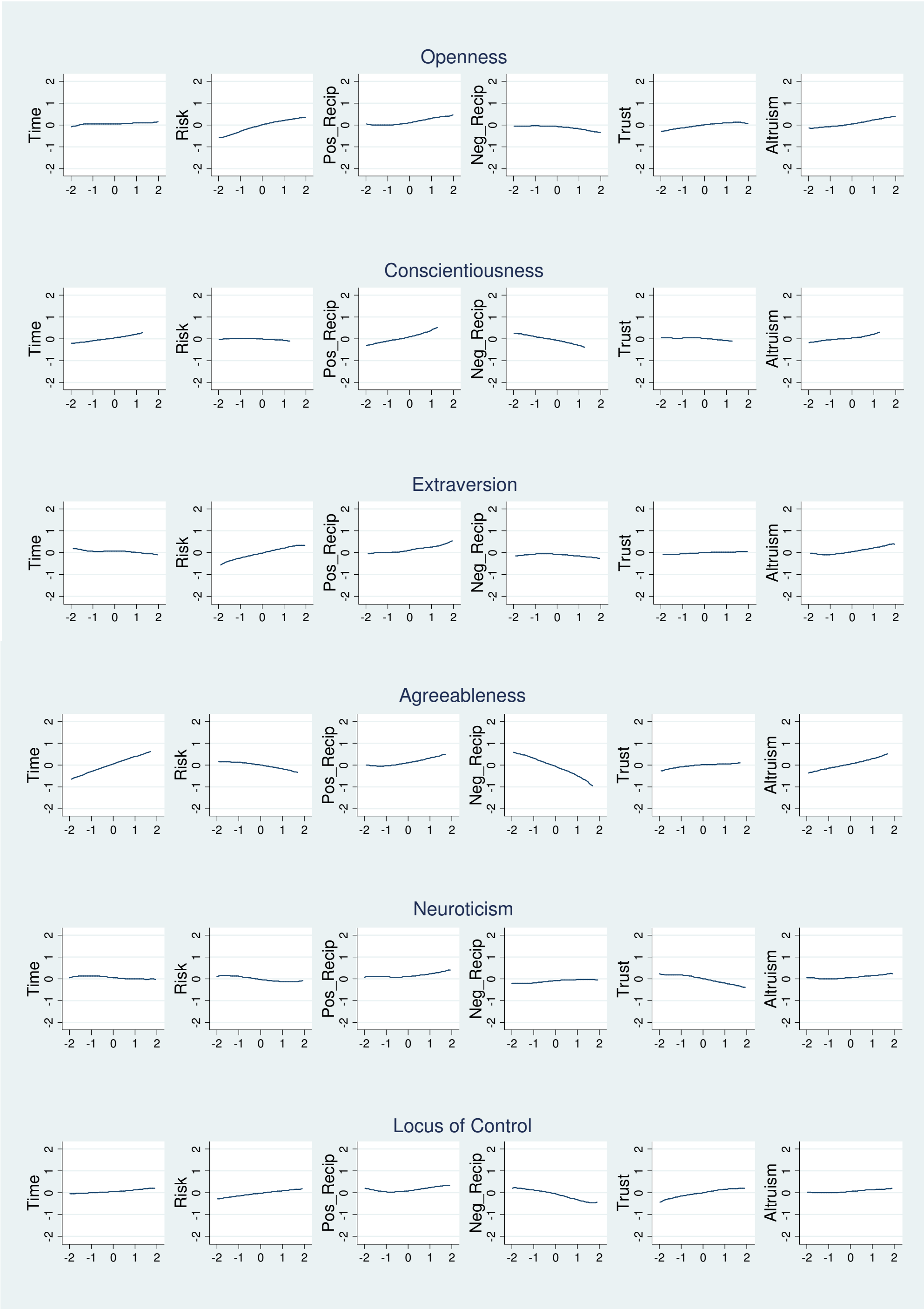
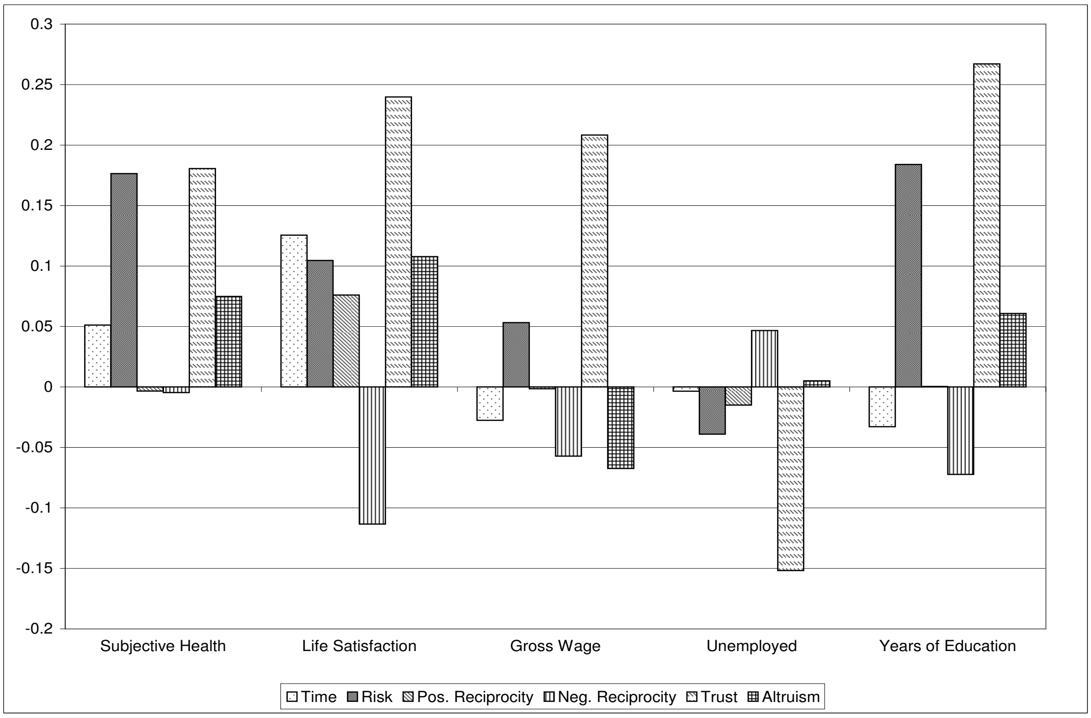
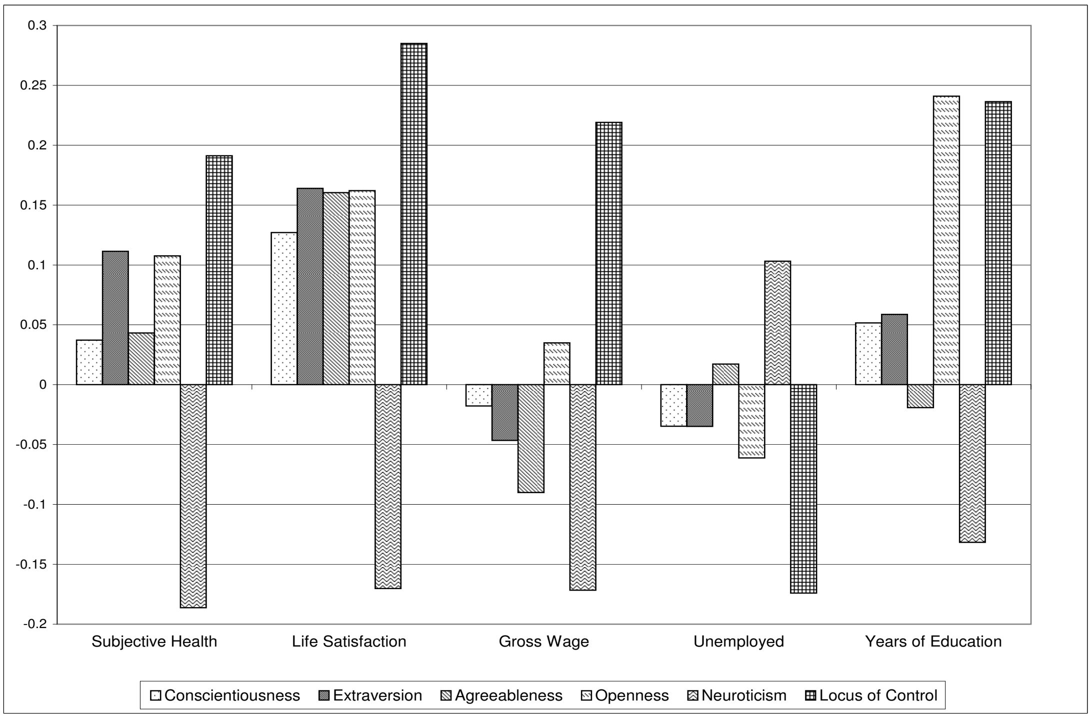


Figure A1.3: Correlation coefficients between preference measures and life outcomes using SOEP data



This graph shows Pearson correlation coefficients between preference measures and life outcomes using SOEP data. Trust always shows the strongest association with life outcomes. More trust and a higher willingness to take risk are always related to better life outcomes, e.g. better health and greater life satisfaction, while negative reciprocity is associated with less life satisfaction and lower wages. The number of observations available varies for the different life outcomes: Subjective Health (14,218 obs.), Life Satisfaction (14,214 obs.), Gross Wage (7,199 obs.), Unemployed (9,095 obs.), Years of Education (13,768 obs.). Gross Wage measures the gross hourly wage.

Figure A1.4: Correlation coefficients between personality measures and life outcomes using SOEP data



This graph shows Pearson correlation coefficients between personality measures and life outcomes using SOEP data. Locus of Control and neuroticism show the strongest associations with life outcomes. A more internal Locus of Control is always related to better outcomes, e.g. better health or more life satisfaction, while a higher degree of neuroticism is associated with lower wages or a higher probability of being unemployed. The number of observations available varies for the different life outcomes: Subjective Health (14,218 obs.), Life Satisfaction (14,214 obs.), Gross Wage (7,199 obs.), Unemployed (9,095 obs.), Years of Education (13,768 obs.). Gross Wage measures the gross hourly wage.

Table A1.6: Outcome regressions: Linear specification

	Subjective Health (OLS)					Subjective Health (o. probit)				
	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC
adj. R^2/pseudo R^2	0.0561	0.0383	0.0688	0.0975	0.1075	0.0220	0.0145	0.0268	0.0388	0.0429
F-Test/LR-Test	170.04	567.35	176.01	140.59	143.72	834.99	550.62	1016.47	1471.22	1627.11
AIC	37833	38094	37641	37201	37043	37139	37415	36960	36515	36361
BIC	37878	38109	37694	37292	37142	37207	37453	37035	36628	36482
	Life Satisfaction (OLS)					Life Satisfaction (o. probit)				
	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC					
adj. R^2/pseudo R^2	0.0899	0.0782	0.0917	0.1342	0.1588	0.0261	0.0219	0.0256	0.0390	0.0467
F-Test/LR-Test	281.88	1206.91	240.08	201.27	224.67	1406.38	1178.16	1376.73	2098.73	2513.61
AIC	55038	55216	55012	54335	53926	52448	52668	52480	51768	51355
BIC	55083	55231	55065	54426	54024	52561	52751	52601	51926	51521
	Gross Wage(OLS)									
	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC	-	-	-	-	-
adj. R^2/pseudo R^2	0.0361	0.0388	0.0456	0.0704	0.0919	-	-	-	-	-
F-Test/LR-Test	54.97	291.20	58.31	50.57	61.71	-	-	-	-	-
AIC	55088	55088	55042	54857	54690	-	-	-	-	-
BIC	55102	55102	55090	54940	54779	-	-	-	-	-
	Unemployed (OLS)					Unemployed (probit)				
	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC
adj. R^2/pseudo R^2	0.0191	0.0331	0.0245	0.0375	0.0547	0.0322	0.0527	0.0412	0.0648	0.0926
F-Test/LR-Test	36.34	312.13	39.05	33.22	44.82	180.12	294.52	230.37	361.89	517.42
AIC	3067	2932	3017	2900	2738	5420	5298	5372	5250	5097
BIC	3110	2946	3067	2986	2830	5463	5312	5422	5336	5189
	Years of Education (OLS)					Years of Education (o. probit)				
	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC					
adj. R^2/pseudo R^2	0.0914	0.0525	0.1061	0.1545	0.1736	0.0209	0.0126	0.0241	0.0359	0.0415
F-Test/LR-Test	277.93	763.89	273.29	229.74	242.03	1355.80	817.10	1563.14	2329.14	2688.38
AIC	65506	66078	65282	64520	64206	63490	64021	63285	62529	62171
BIC	65551	66093	65335	64610	64304	63641	64141	63443	62724	62375

The outcome variables are regressed on the indicated personality and preference measures. For OLS models we calculate R^2, for ordinal models we calculate pseudo R^2. Joint significance of all coefficients is tested using the F-Test (OLS) and the LR-Test (ordinal models). All F- and LR-Tests are significant at the 1% level. Concerning the Akaike information criterion (AIC) and Bayesian information criterion (BIC), the smallest value for each outcome regression is underlined. Note that the full model (including Big5, LoC and Pref) is always chosen by both information criteria. The number of observations available varies for the different life outcomes: Subjective Health (14,218 obs.), Life Satisfaction (14,214 obs.), Gross Wage (7,199 obs.), Unemployed (9,095 obs.), Years of Education (13,768 obs.). Gross Wage measures the gross hourly wage.

Table A1.7: Outcome regressions: Flexible specification

	Subjective Health (OLS)					Subjective Health (o. probit)				
	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC
adj. R^2/pseudo R^2	.0632	.0388	.0714	.1054	.1165	.0251	.0146	.0282	.0435	.0483
F-Test/LR-Test	48.99	288.17	41.48	22.75	21.83	952.98	555.19	1068.56	1651.38	1834.03
AIC	37740	38088	37623	37142	36977	37051	37413	36949	36467	36310
BIC	37899	38110	37834	37732	37665	37232	37458	37184	37079	37021
	Life Satisfaction (OLS)					Life Satisfaction (o. probit)				
	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC
adj. R^2/pseudo R^2	.0948	.0783	.0948	.1397	.1659	.0278	.0219	.0273	.0422	.0505
F-Test/LR-Test	75.47	605.45	56.12	30.967	32.41	1493.78	1178.45	1470.26	2273.51	2715.76
AIC	54976	55214	54984	54311	53884	52391	52670	52428	51725	51309
BIC	55135	55237	55196	54901	54572	52617	52761	52708	52383	52065
	Gross Wage(OLS)									
	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC	-	-	-	-	-
adj. R^2/pseudo R^2	.0382	.0387	.0527	.0797	.1039	-	-	-	-	-
F-Test/LR-Test	15.30	145.74	15.84	9.092	10.27	-	-	-	-	-
AIC	55111	55090	55009	54851	54672	-	-	-	-	-
BIC	55256	55111	55202	55388	55298	-	-	-	-	-
	Unemployed (OLS)					Unemployed (probit)				
	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC
adj. R^2/pseudo R^2	.0212	.0385	.0291	.0463	.0705	.0357	.0539	.0498	.0852	.1166
F-Test/LR-Test	10.87	183.13	11.11	6.73	8.66	199.54	301.02	278.38	475.96	651.83
AIC	3062	2882	2995	2882	2662	5431	5294	5366	5268	5118
BIC	3211	2903	3194	3437	3309	5580	5314	5565	5823	5766
	Years of Education (OLS)					Years of Education (o. probit)				
	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC	Big5	LoC	Pref	Big5-Pref	Big5-Pref-LoC
adj. R^2/pseudo R^2	.1043	.0525	.1200	.1771	.1982	.0243	.0126	.0281	.0433	.0497
F-Test/LR-Test	81.13	382.50	70.55	39.48	38.81	1575.60	817.25	1819.82	2808.59	3223.85
AIC	65324	66079	65087	64213	63869	63300	64023	63070	62181	61792
BIC	65482	66102	65297	64800	64554	63564	64151	63386	62874	62583

The outcome variables are regressed on the indicated personality and preference measures. The difference to the linear specification is that the model includes squares of all variables as well as all cross-products. Cross-products are also calculated between concepts in case more than one concept is included, e.g., in the Big5-Pref case, we also include (among others) the cross term neuroticicsm*risk. For OLS models we calculate R^2, for ordinal models we calculate pseudo R^2. Joint significance of all coefficients is tested using the F-Test (OLS) and the LR-Test (ordinal models). All F- and LR-Tests are significant at the 1% level. Concerning the Akaike information criterion (AIC) and Bayesian information criterion (BIC), the smallest value for each outcome regression is underlined. Note that the full model (including Big5, LoC and Pref) is chosen by both information criteria in nearly all cases; only for cross wage and unemployment the BIC indicates to use the model with only LoC and LoC^2 included. The number of observations available varies for the different life outcomes: Subjective Health (14,218 obs.), Life Satisfaction (14,214 obs.), Gross Wage (7,199 obs.), Unemployed (9,095 obs.), Years of Education (13,768 obs.).

A2 Appendix to Chapter 2

Table A2.1: Relations of traits to cog. and noncog. factors

	Cog		NC-L		NC-E		NC-R		NC-B	
Cons	-0.042	-0.069**	0.138***	0.122***	0.068**	0.053	0.067**	0.038	0.165***	0.148***
	(0.030)	(0.031)	(0.028)	(0.029)	(0.032)	(0.033)	(0.031)	(0.032)	(0.041)	(0.043)
Agree	0.102***	0.108***	-0.086***	-0.087***	0.091***	0.086***	-0.162***	-0.153***	-0.011	0.009
	(0.030)	(0.030)	(0.028)	(0.028)	(0.032)	(0.032)	(0.031)	(0.031)	(0.041)	(0.041)
Neuro	-0.274***	-0.279***	-0.433***	-0.433***	-0.033	-0.028	-0.082***	-0.090***	0.100**	0.083**
	(0.029)	(0.029)	(0.027)	(0.027)	(0.030)	(0.030)	(0.030)	(0.030)	(0.040)	(0.040)
Open	0.587***	0.588***	-0.011	-0.016	0.250***	0.240***	0.080*	0.083*	0.012	0.036
	(0.045)	(0.045)	(0.043)	(0.043)	(0.048)	(0.048)	(0.047)	(0.047)	(0.064)	(0.064)
Extrav	-0.473***	-0.453***	0.015	0.015	-0.068	-0.077	0.018	0.047	-0.164***	-0.141**
	(0.044)	(0.045)	(0.042)	(0.042)	(0.047)	(0.047)	(0.046)	(0.047)	(0.063)	(0.063)
Time		0.077***		0.054**		0.066**		0.074***		0.014
		(0.027)		(0.025)		(0.028)		(0.028)		(0.037)
Risk		-0.034		0.025		0.067**		-0.069**		-0.127***
		(0.027)		(0.026)		(0.029)		(0.028)		(0.038)
N	1382	1382	1382	1382	1382	1382	1382	1382	758	758
R^2	0.128	0.135	0.249	0.251	0.043	0.050	0.075	0.085	0.057	0.072

Notes: Table shows regressions of Cognition and different Non-cognitive constructs on the Big-5 Personality traits, discount rate, and risk preference. NC-L is based on the Rotter's Locus of control. NC-E is based on engagement behavior, NC-R is based on self-reported relationships, and NC-B is based on self reported risky behaviors. Standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2.2: Model comparison: In college (marginal effects)

	NC-L	NC-E	NC-R	NC-B	Pref-1	Pref-2	Comb-L	Comb-E	Comb-R	Comb-B
IQ	0.139***	0.147***	0.146***	0.146***	0.158***		0.149***	0.159***	0.160***	0.156***
	(0.034)	(0.033)	(0.033)	(0.033)	(0.036)		(0.037)	(0.036)	(0.036)	(0.036)
Noncog	0.034	-0.008	-0.000	0.013			0.037	-0.016	-0.036	-0.016
	(0.036)	(0.038)	(0.038)	(0.038)			(0.043)	(0.038)	(0.039)	(0.039)
Cons.					0.109**	0.116***	0.102**	0.111***	0.113***	0.113***
					(0.043)	(0.045)	(0.043)	(0.043)	(0.043)	(0.044)
Agree.					-0.029	0.023	-0.024	-0.025	-0.036	-0.027
					(0.043)	(0.044)	(0.043)	(0.044)	(0.044)	(0.044)
Neuro.					0.078*	0.033	0.091*	0.079*	0.076*	0.077*
					(0.045)	(0.046)	(0.048)	(0.045)	(0.045)	(0.045)
Open.					-0.039	0.068	-0.025	-0.036	-0.038	-0.037
					(0.066)	(0.063)	(0.067)	(0.066)	(0.066)	(0.066)
Extra.					0.076	-0.003	0.066	0.075	0.081	0.072
					(0.063)	(0.063)	(0.064)	(0.063)	(0.063)	(0.064)
Risk					-0.044	-0.047	-0.047	-0.042	-0.042	-0.047
					(0.037)	(0.038)	(0.037)	(0.037)	(0.037)	(0.038)
Time					-0.031	-0.034	-0.039	-0.029	-0.030	-0.032
					(0.039)	(0.040)	(0.040)	(0.039)	(0.039)	(0.039)
Observations	177	177	177	177	177	177	177	177	177	177
Pseudo R^2	0.099	0.096	0.095	0.096	0.152	0.086	0.155	0.153	0.155	0.153

Notes: Table shows Probit estimations of college enrollment on one of the four constructed 2-factor models, our two preferred models, or combined models. The displayed coefficients are average marginal effects. NC-L is based on the Rotter's Locus of control. NC-E is based on engagement behavior, NC-R is based on self-reported relationships, and NC-B is based on self reported risky behaviors. All estimated probit models include the following controls: gender, urban status, and residence in Eastern Germany. Standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2.3: Model comparison: GPA, including additional controls: Eduation of parents

	NC-L	NC-E	NC-R	NC-B	Pref-1	Pref-2	Comb-L	Comb-E	Comb-R	Comb-B
IQ	0.203***	0.208***	0.206***	0.228***	0.192***		0.196***	0.191***	0.190***	0.203***
	(0.022)	(0.021)	(0.021)	(0.028)	(0.021)		(0.022)	(0.021)	(0.021)	(0.029)
Noncog	0.036*	0.030	0.057***	0.083***			-0.018	0.019	0.025	0.056**
	(0.020)	(0.019)	(0.019)	(0.026)			(0.022)	(0.019)	(0.019)	(0.026)
Cons.					0.181***	0.173***	0.184***	0.180***	0.179***	0.179***
					(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.030)
Agree.					0.057***	0.079***	0.055**	0.055**	0.061***	0.071**
					(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.028)
Neuro.					-0.026	-0.071***	-0.033	-0.026	-0.025	-0.029
					(0.022)	(0.022)	(0.024)	(0.022)	(0.022)	(0.029)
Open.					0.017	0.120***	0.013	0.013	0.016	0.002
					(0.036)	(0.036)	(0.037)	(0.037)	(0.036)	(0.047)
Extra.					-0.021	-0.100***	-0.018	-0.020	-0.023	0.016
					(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.046)
Risk					-0.057***	-0.063***	-0.056***	-0.059***	-0.055***	-0.054**
					(0.020)	(0.021)	(0.020)	(0.020)	(0.020)	(0.027)
Time					0.028	0.039*	0.028	0.027	0.026	0.040
					(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.026)
Observations	1217	1217	1217	715	1217	1217	1217	1217	1217	715
R^2	0.162	0.162	0.166	0.167	0.229	0.177	0.229	0.229	0.230	0.238

Notes: Table shows regressions of GPA on one of the four constructed 2-factor models, our two preferred models, or combined models. NC-L is based on the Rotter's Locus of control. NC-E is based on engagement behavior, NC-R is based on self-reported relationships, and NC-B is based on self reported risky behaviors. All estimated OLS models include the following controls: parent's education, gender, urban status, residence in Eastern Germany and the education tier in which the grade was received. Standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2.4: Model comparison: In college, including additional controls: Eduation of parents (marginal effects)

	NC-L	NC-E	NC-R	NC-B	Pref-1	Pref-2	Comb-L	Comb-E	Comb-R	Comb-B
IQ	0.130***	0.135***	0.133***	0.134***	0.152***		0.151***	0.152***	0.154***	0.148***
	(0.039)	(0.039)	(0.039)	(0.039)	(0.041)		(0.042)	(0.041)	(0.041)	(0.041)
Noncog	0.016	-0.018	-0.008	0.008			0.005	-0.023	-0.044	-0.024
	(0.036)	(0.037)	(0.038)	(0.038)			(0.045)	(0.037)	(0.038)	(0.040)
Cons.					0.116***	0.125***	0.115***	0.119***	0.122***	0.123***
					(0.043)	(0.045)	(0.044)	(0.043)	(0.043)	(0.044)
Agree.					-0.026	0.017	-0.025	-0.021	-0.035	-0.023
					(0.043)	(0.043)	(0.043)	(0.043)	(0.044)	(0.043)
Neuro.					0.077*	0.041	0.079*	0.079*	0.076*	0.076*
					(0.044)	(0.045)	(0.047)	(0.044)	(0.044)	(0.044)
Open.					-0.066	0.014	-0.064	-0.061	-0.066	-0.063
					(0.064)	(0.063)	(0.067)	(0.065)	(0.065)	(0.064)
Extra.					0.088	0.026	0.087	0.086	0.095	0.081
					(0.062)	(0.062)	(0.063)	(0.062)	(0.062)	(0.063)
Risk					-0.047	-0.045	-0.047	-0.044	-0.045	-0.052
					(0.041)	(0.042)	(0.041)	(0.041)	(0.041)	(0.041)
Time					-0.023	-0.022	-0.024	-0.020	-0.020	-0.024
					(0.037)	(0.038)	(0.038)	(0.037)	(0.037)	(0.037)
Observations	173	173	173	173	173	173	173	173	173	173
Pseudo R^2	0.120	0.120	0.119	0.119	0.176	0.126	0.176	0.178	0.181	0.178

Notes: Table shows Probit estimations of college enrollment on one of the four constructed 2-factor models, our two preferred models, or combined models. The displayed coefficients are average marginal effects. NC-L is based on the Rotter's Locus of control. NC-E is based on engagement behavior, NC-R is based on self-reported relationships, and NC-B is based on self reported risky behaviors. All estimated probit models include the following controls: parent's education, gender, urban status, and residence in Eastern Germany. Standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A3 Appendix to Chapter 3

	Rotated principal components		
	High quality time	**Medium quality time**	**Low quality time**
How many times in the last 14 days have you or the main caregiver done the following activities together with your child?			
Singing children's songs with or to the child	**0.4443**	-0.0034	-0.1241
Reading or telling stories	**0.5787**	-0.0178	-0.1144
Looking at picture books	**0.5576**	-0.0496	-0.0090
Painting or doing arts and crafts	0.3475	0.1182	0.3454
Taking walks outdoors	0.1022	**0.4269**	0.1194
Going to the playground	0.011	**0.5464**	0.0019
Visiting other families with children	-0.0336	**0.5602**	-0.2454
Going shopping with the child	-0.1521	**0.4375**	0.0513
Watching television or videos with the child	-0.0799	-0.085	**0.8879**

Table A3.1: Principal component analysis concerning the quality of the parent-child interaction (age 2-3 years). Source: SOEP (2012); $N = 552$; Mothers are asked how many times in the last 14 days she, or the main caregiver, has done particular activities together with their child. Using the answers concerning all nine potential activities we performed a principal component analysis (rotation method: Oblique promax (power = 3), resulting in three components according to Kaiser Criterion (Eigenvalue > 1). The first component reflects activities, which involve face-to-face contact and a high degree of interaction between mother and child such as reading or telling children's stories or singing children's songs with the child (high quality time). The second component reflects activities with a medium degree of interaction and less direct contact such as going shopping or visiting other families with the child (medium quality time). The third component represents watching TV or videos (low quality time).

	Breastfed	Duration of BF (if BF >0)
	Binary (yes=1, no=0)	In months
	Probit	OLS
	(1)	(2)
Parent-child interaction		
Component high quality time (age 2-3)	0.019*	0.852***
	(0.010)	(0.170)
Component medium quality time (age 2-3)	-0.011	-0.396*
	(0.011)	(0.219)
Component low quality time (age 2-3)	-0.006	-0.496*
	(0.014)	(0.283)
Physical health problems of mother	-0.113***	-0.435
(last third of pregnancy and 3 months after birth)	(0.028)	(0.626)
Socio-economic status		
College degree mother	0.141***	1.098
	(0.047)	(0.715)
Log net household income	-0.009	0.529
	(0.029)	(0.586)
Constant		2.973
		(4.602)
Cohort dummies	Yes	Yes
Wald-tests:		
- all parent-child interaction = 0	$\chi^2 = 4.00$	$F = 9.57^{***}$
- all socio-economic status = 0	$\chi^2 = 9.06^{**}$	$F = 2.12$
Observations	552	484
(Pseudo) R-squared	0.120	0.074

Table A3.2: Determinants of breastfeeding duration. Source: SOEP (2012). The displayed coefficients are average marginal effects. For estimation of the components of parent-child interaction, see Table A3.1. Physical health problems of mother is a dummy indicating rather bad or very bad health in last third of pregnancy or the first three months after birth. College degree mother is a dummy variable indicating whether mother holds a university or technical college degree. Net household income is the self-reported net household income. For Wald-tests concerning the OLS (Probit) estimations F- (χ^2-) values are displayed. Clustered standard errors (at household level) in parentheses; ***, **, * indicate significance at 1-, 5-, and 10-percent level, respectively.

Correlations between breastfeeding duration (BF > 0) and	**Spearman's rho**	**P-value**
HOME inventory (at age: 3 months)[a]	0.126	0.033
HOME inventory (at age: 2 years)[a]	0.177	0.003
HOME inventory (at age: 4.5 years)[a]	0.175	0.003
Importance of having children for mother[b]	0.111	0.015
Life satisfaction of the mother (in the year of birth of the child)[b]	0.099	0.030

Table A3.3: Correlations of breastfeeding duration and other variables reflecting the quality of early life circumstances. Sources: [a] Mannheim Study of Children at Risk (MARS) (Blomeyer et al., 2009) ($N = 384$) and [b] SOEP (2012) ($N = 484$). We acknowledge provision of correlations concerning HOME Inventory by Karsten Reuß. Displayed coefficients are Spearman rank correlation coefficients. Home Observation for Measurement of the Environment (HOME) (Bradley and Caldwell, 1981; Blomeyer et al., 2009) is a 26 item rating. Importance of having children is measured on a 4-point scale in the year 2008 when all children were already born. Life satisfaction of the mother is measured in the year of birth of the child and is measured on an 11-point Likert scale.

	Time (0/1) Probit		Risk (standardized) OLS		Altruism (0/1) Probit	
	(1)	(2)	(3)	(4)	(5)	(6)
Breastfeeding						
Duration of breastfeeding	0.025***	0.024***	-0.032**	-0.033*	0.016**	0.021***
(in months)	(0.008)	(0.008)	(0.016)	(0.020)	(0.007)	(0.006)
Child's characteristics						
Age (in months)		0.004		0.014		-0.002
		(0.006)		(0.021)		(0.009)
Dummy male		0.014		0.617***		-0.087
		(0.056)		(0.196)		(0.070)
Height (in 10 cm)		0.028		-0.288		0.203***
		(0.040)		(0.200)		(0.065)
Intelligence (standardized)		0.038		0.030		0.013
		(0.031)		(0.150)		(0.048)
Socio-economic environment						
College degree mother		-0.085		-0.101		0.084
		(0.082)		(0.237)		(0.071)
Log net household income		-0.065		-0.015		-0.009
		(0.069)		(0.281)		(0.065)
Dummy older siblings		-0.017		-0.054		0.037
		(0.071)		(0.209)		(0.071)
Dummy younger siblings		0.084		0.060		0.034
		(0.073)		(0.239)		(0.064)
Age of mother (in years)		0.006		0.004		-0.009
		-0.085		-0.101		0.084
Personality/preferences/ IQ of mother						
Openness to experience		-0.007		0.149		-0.068
		(0.029)		(0.131)		(0.043)
Conscientiousness		0.020		0.118		0.012
		(0.030)		(0.104)		(0.036)
Extraversion		0.024		-0.017		0.059**
		(0.027)		(0.090)		(0.029)
Agreeableness		0.031		-0.066		0.037
		(0.029)		(0.117)		(0.036)
Neuroticism		0.043		-0.137		0.058*
		(0.035)		(0.103)		(0.034)
Intelligence		-0.027		-0.194		0.014
		(0.028)		(0.125)		(0.040)
Time preference		0.071**				
		(0.030)				
Risk preference				-0.113		
				(0.099)		
Altruism						0.136
						(0.103)
Task specific controls	no	yes	no	no	no	yes
Observations	194	194	108	108	100	100
(Pseudo) R-squared	0.047	0.179	0.025	0.183	0.058	0.286

Table A3.4: The effect of quality of early life circumstances on preschool children's preferences (Data Set 1). Displayed coefficients are average marginal effects with respective preference as dependent variable. In the estimations we use age and age squared of the child as explanatory variables. The combined intelligence measure of the child is the standardized score of standardized fluid and crystallized intelligence. The variable "net monthly household income" refers to the current monthly income in €of all household members, net of taxes and benefits. For less than 20 percent of respondents, income was only reported in intervals (<750; 750 - 1,500; 1,500 - 2,500; 2,500 - 3,500; 3,500 - 5,000; >5,000 Euros). In these cases we used the interval midpoints (7,500 in case of income exceeding 5,000). All personality, preference and IQ measures of the mother are standardized, the only exception is altruism which is a dummy indicating selecting the altruistic distribution. Intelligence of the mother is measured by the number of correct answers in a symbol-digit-test. Time preference of mothers' is the (reversed) switching row in the time preference choice task, risk preference is the certainty equivalent in the lottery task. Specific controls in column (2) are dummies indicating elapsed time since the last bigger meal and in column (6) a dummy indicating if the matched child is from the same kindergarten. To receive comparable results we exclude observations with missing values in the covariates from all regressions. Robust standard errors in parentheses. ***, **, * indicate significance at 1-, 5-, and 10-percent level, respectively.

	Time (std.)		Risk (std.)		Altruism (std.)	
	OLS		OLS		OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Breastfeeding						
Duration of breastfeeding	0.042**	0.038**	-0.036*	-0.040**	0.048***	0.045***
(in months)	(0.016)	(0.016)	(0.020)	(0.019)	(0.016)	(0.015)
Child's characteristics						
Age (in months)		-0.030		0.030		0.056*
		(0.034)		(0.034)		(0.034)
Dummy male		0.155		0.033		-0.036
		(0.162)		(0.160)		(0.161)
High school math grade		-0.140**		-0.038		-0.079
(low is better)		(0.071)		(0.077)		(0.069)
Socio-economic environment						
Occupation of father (dummies)	No	Yes++	No	Yes++	No	Yes+
Occupation of mother (dummies)	No	Yes+	No	Yes+	No	Yes
Observations	175	175	175	175	175	175
R-squared	0.028	0.222	0.021	0.210	0.037	0.165

Table A3.5: The effect of quality of early life circumstances on young adults' preferences (Data Set 2). Displayed coefficients are marginal effects, with respective standardized preference measure as dependent variable and robust standard errors in parentheses. In the estimations we use age and age squared of the young adult as explanatory variables. High school math grade serves as a proxy for IQ and is coded in the typical German 6-point grading system where lower values indicate better performance. Socio-economic environment is controlled for by including dummies indicating occupation of father and mother. ***, **, * indicate significance at 1-, 5-, and 10-percent level, respectively. +++, ++, + indicate significance at 1-, 5-, and 10-percent level, of Wald-tests testing the hypothesis that all coefficients of the respective category are zero.

Age: 11-17	Current smoker		Ever drunk alcohol		BMI	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Binary		Binary		Continuous	
Type of estimation	Probit		Probit		OLS	
Breastfeeding duration	-0.006***	-0.003**	-0.006***	-0.004**	-0.050***	-0.026**
	(0.002)	(0.001)	(0.002)	(0.002)	(0.013)	(0.012)
Dummy male		0.000		-0.013		-0.196*
		(0.011)		(0.014)		(0.112)
Age (in years)		0.062***		0.081***		0.613***
		(0.003)		(0.003)		(0.028)
College degree mother		-0.034**		-0.003		-0.510***
		(0.014)		(0.019)		(0.127)
Log net HH income		-0.048***		0.029**		-0.612***
		(0.010)		(0.014)		(0.118)
Observations	4,395	4,395	4,123	4,123	4,416	4,416
(Pseudo) R-squared	0.005	0.200	0.002	0.126	0.003	0.109

Table A3.6: The effect of quality of early life circumstances on health-related behaviors and outcomes. Source: KiGGS (2008). Displayed coefficients are average marginal effects with robust standard errors in parentheses. In the estimations we use age and age squared as explanatory variables. ***, **, * indicate significance at 1-, 5-, and 10-percent level, respectively.

	Time (11-p. scale) (1)	Risk (11-p. scale) (2)	Altruism (4-p. scale) (3)
Av. duration (in months) x share	0.252	-0.291**	0.159***
	(0.533)	(0.121)	(0.044)
Share not breastfed (in %)	0.045*	-0.002	0.002
	(0.018)	(0.009)	(0.003)
Age	-0.003	-0.080***	0.016**
	(0.077)	(0.012)	(0.006)
Constant	3.916	7.806***	2.562***
	(2.778)	(0.644)	(0.230)
Observations	8	47	14
R-squared	0.783	0.395	0.600
Adj. R-squared	0.620	0.353	0.479

Table A3.7: A cohort level analysis of the effect of quality of early life circumstances on preferences. Source: Nestlé (see Figure A3.2) for breastfeeding durations and shares, and SOEP (2012) for preference measures. One preference observation reflects the average value of a preference for one birth cohort at a given year. For risk and altruism we estimate a pooled OLS and show clustered standard errors (at birth cohort levels) in parentheses, for time preference data are only cross sectional. The panel structure of our data set concerning risk and altruism enables us to disentangle the breastfeeding duration effect from an age effect. For interpretations and details see Methods. ***, **, * indicate significance at 1-, 5-, and 10-percent level, respectively.

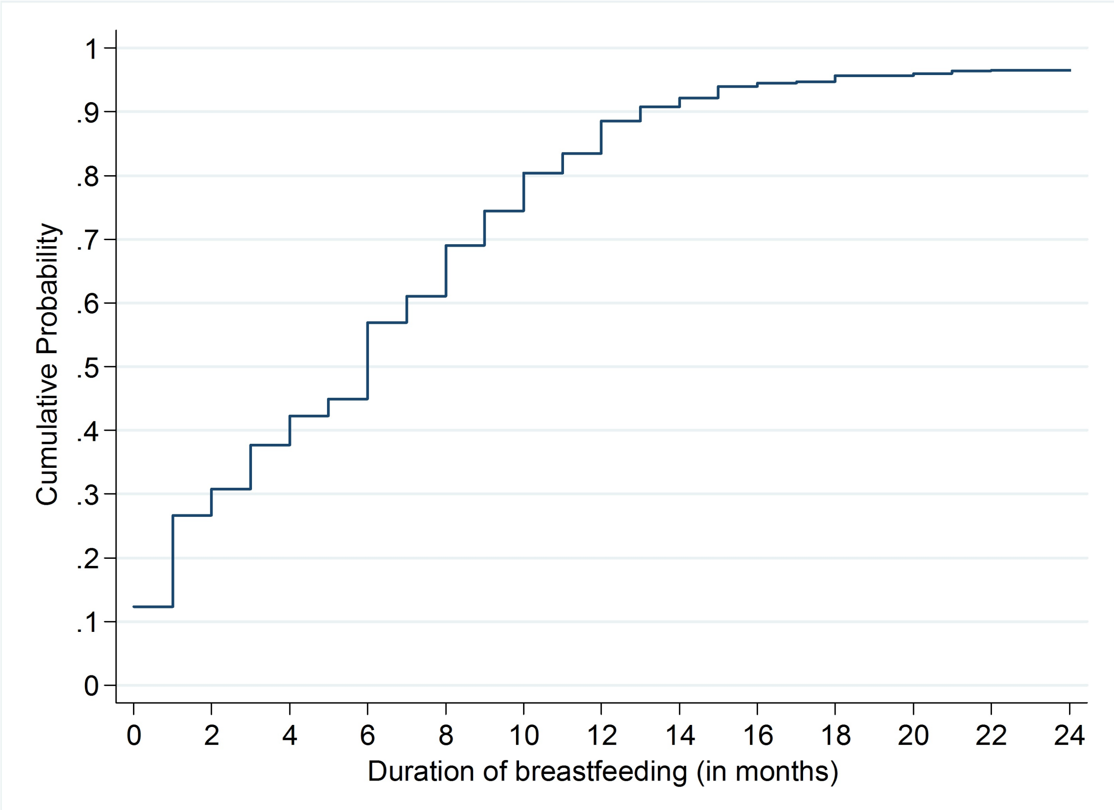


Figure A3.1: The distribution of breastfeeding durations in Germany for birth cohorts 2004-2007. Source: SOEP (2012). $N = 552$.

Figure A3.2: The development of breastfeeding durations and shares in Germany for birth cohorts 1976-1992. Source: Nestlé. $N = 250$ per cohort.

A4 Appendix to Chapter 4

Details on study design

Recruitment of sample and randomization

Using official registry data, we received more than 95% of the addresses of families living in the German cities Bonn and Cologne who had children of age seven to nine when the study started (October 2011). Offers to take part in the study were sent by mail to all families with children born between 01.09.2003 and 31.08.2004 and one third of the families with children born between 01.09.2002 and 31.08.2003. In summer 2011, families were informed about the study which included two waves of interviews, focused on child development and offered the possibility to take part in a mentoring program via postal mail. If interested in the study, we asked them to send back a short questionnaire on socio-economic characteristics of the household and to sign an unbinding letter of intent to take part in the interviews and potentially the mentoring program. Using answers in the survey, we categorized families according to socio-economic status (SES). Target group families (low SES) meet at least one of three criteria; their exact definitions are as follows:

1. Low income: Equivalence income of the household is lower than the 1065 Euro, i.e., the 30% quantile of the German income distribution.
2. Low education: Both mother and father of the child have at most secondary education, i.e., are not qualified for university studies.
3. Single parents: The single parent is not living together with a partner.

We invited all families who belonged to the target group to take part in the interviews. To be eligible for treatment, they must have taken part in the first wave of interviews and experiments (fall 2011) and given written consent to transmit their addresses to the organization running the mentoring program. Out of 590 eligible families, 212 were stratified randomly selected and form the intention-to-treat (ITT) group. The stratification considered 14 subgroups resulting from the combination of local (Cologne or Bonn) and target (low income and/or low education and/or single parent) criteria. Stratification was used to ensure two things: First, a proportional representation of all criteria combinations in the ITT group and second, an adaption to the local supply of mentors. 151 out of the 212 ITT group children were actually treated and met their mentor at least once. 83.7% (494 out of 590) eligible families took part in the second wave of interviews and experiments in the beginning of 2013 and build our core sample. The "lost to follow-up rates" do not differ in treatment and control group ($p = 0.563, N = 590$, two-sided test of proportions).

As a second control group we also invited 150 randomly chosen high SES families (out of those who also answered to our study information) to take part in the study. To ensure comparability we also asked them to given written consent to transmit their addresses to the organization running the mentoring program. 122 took part in the first wave and gave the written consent. Out of them 113 took part in the second wave of interviews and experiments.

The intervention: A mentoring program

The intervention we randomly implemented is a non-profit mentoring program called "Balu und Du" (German for "Baloo and You"). In this program, children get a mentor by their side for a period of one year. The mentors are mainly university students aged between 18 and 30 who wish to volunteer alongside their studies. Ideally, children meet with their mentor once a week and, together, engage in versatile activities such as visiting the zoo, museum, or park, cooking, ice skating, visiting the playground or just having a conversation. The pedagogical idea of the program is to enrich children's life circumstances and to extend their horizon via joint activities with a new contact person. An important feature of the program is the mere circumstance that there is a further person who is responsive to a child's individual needs and interests. The program fosters the acquisition of new skills on an informal basis.

The mentoring program is embedded in a professional structure. Mentors administer an online diary in which they report the activities they have engaged in and potential problems of the mentor-child relationship on a weekly basis. Program coordinators read and comment these diaries, and provide support. The coordinators are trained and paid professionals in education science or psychology who supervise, coach and advice mentors on a part or full-time basis. They also organize biweekly monitoring meetings in which mentors can discuss potential problems and receive suggestions for activities with the mentored child. To date, the mentoring program “Balu und Du” has arranged and supervised more than 5,500 mentor-child relationships in over 50 different locations in Germany. The program has been honored with numerous awards. More details about the mentoring program can be found on www.balu-and-du.de. For comparisons to other mentoring or school-based programs see the overview studies of Rodríguez-Planas (2012) and Heckman and Kautz (2014).

The setting of interviews and experiments

The families visited a central location in Bonn or Cologne, Germany, respectively. They participated in interviews and experiments that were conducted by trained university students (mostly graduates) of psychology or education science. The interviews and experiments were conducted according to a detailed protocol (see below). In total, the interviews lasted about one hour. Children were paid and incentivized using an experimental currency called “stars”. At the end of the interview, children could exchange their stars into toys. As displayed in Fig. A4.1, toys were arranged in four categories which visibly increased in objective value and subjective attractiveness to children.

During the experiments, children knew that more stars would result in the option to choose a toy from a higher category. We ensured that each additional star that would not result in a higher category still had an extra value to the children by converting these additional stars into “Lego” bricks. While their children participated in the experiments, mothers filled out a comprehensive questionnaire covering the following topics:

- Basic information about the child, e.g., name, age, etc.
- Mother assessments of personality and attitudes of the child

Figure A4.1: Toys arranged in four categories

- Socio-economic background of the family
- Personality, preferences, and attitudes of the mother

Detailed description of experiments and questionnaire measures

Altruism: Incentivized dictator games

Altruism was elicited using three versions of dictator games in which children were in the role of the decision maker. In particular, we played one binary dictator game and two continuous versions of dictator games with varying receivers to learn more about the distributions of altruistic giving.

In the binary dictator game, children had to decide between two possible allocations of two stars between themselves and another unknown child from the same city of similar age. Either both receiver and decision maker received one star (1,1) or the decision maker got two stars, while the receiver got zero stars (2,0). Both possible allocations were physically shown to the children and interviewers checked whether children understood the implications of each allocation.

In both of the continuous versions of the dictator game, interviewers showed children two paper bags, one belonging to the child itself and the other belonging to another child, the receiver. In the first version of the game, children were told that the receiver is of the same age as they are, lives in a city nearby, but is unknown to them, and has no relation to the interviewer. In the second version of the game, children were told that the receiver is of the same age, but lives in an African country and cannot live with his parents since they are either ill or dead. In both versions, children were endowed with 6 stars and could choose how to distribute the 6 stars among the two bags. After children had distributed the stars among the two bags, the interviewer checked that they understood how many stars they and the other child received. Only in case they did not understand the resulting allocation, the rules were explained again and children had the opportunity to alter their decision.

Assuming that children's behavior in the three dictator games is a function of their true degree of altruism and some random error component, we can reduce measurement error by using the average share of giving as our measure of altruism.

Trust: Questionnaire answers of the child

In the child questionnaire, children had to rate a series of statements in terms of how much they agreed with each of the statements. The statements were read out aloud by the interviewer and children indicated on a five point scale ranging from "totally correct" to "totally incorrect" how they rated the statements. As displayed in Fig. A2, the scale was printed on an extra paper sheet and additionally visualized by, e.g., using an "X" for "Totally incorrect" and a "check mark" for "Totally correct". The interviewer explained the procedure using a simple example item (I like Spaghetti).

We used three items to infer children's degree of trust. Our measure is based on the three validated trust questions used in the German Socio Economic Panel Study (SOEP) (Fehr et al., 2002). We adapted and reformulated these items in order to be appropriate for children in the age range under study. In particular, the statements are "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".

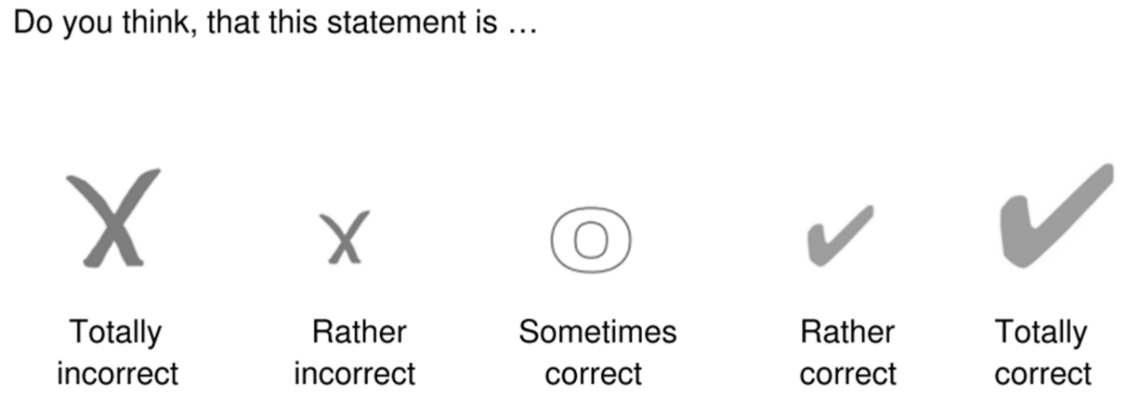


Figure A4.2: Rating scale for the child questionnaire

Other-regarding behavior: Questionnaire answers of the mother

We asked the mother to rate her child's other-regarding behavior in everyday life using the seven items of the Strength and Difficulties Questionnaire (SDQ) which directly refer to children's behavior regarding others. The SDQ is a well-established behavioral screening questionnaire. In the version of the SDQ that we used, parents had to rate statements about their child on a seven point Likert scale from "does not apply at all" (1) to "applies completely" (7). The seven statements of the SDQ that refer to other-regarding behavior are "My child. . . " "Shares readily with other children (treats, toy, pencils etc.)", "Is helpful if someone is hurt, upset or feeling ill", "Often fights with other children or bullies them" (reversed), "Gets along better with adults than with other children" (reversed), "Is generally liked by other children", "Is kind to younger children", "Often volunteers to help others (parents, teachers or other children)".

Prosociality of mothers and mentors

In order to obtain measure of prosociality of mothers and mentors, respectively, we proceeded as similar as possible as for the children: we performed a principal component analysis using standardized measures of altruism, trust and other-regarding behavior resulting in one component according to the Kaiser Criterion (Eigenvalue > 1). All measures are elicited using standardized and validated questionnaire items.

Altruism was measured using the question "How do you assess your willingness to share without expecting anything in return, e.g. your willingness to give to charity?" (Falk et al., 2011). Trust was measured by the item "In general, one can trust people" (Fehr et al., 2002). In both cases answers are given on an 11-point Likert scale. Other-regarding behavior was measured using the Big Five dimension Agreeableness in form of a 3-items short version (Becker et al., 2012).

Additional Tables

	Prosociality (standardized)				
	(1)	(2)	(3)	(4)	(5)
Treatment Dummy	0.305***	0.164	0.318***	0.362***	0.223
	(0.092)	(0.122)	(0.122)	(0.125)	(0.245)
Prosociality of mother	0.244***				0.246***
	(0.052)				(0.052)
Prosociality of mother x Treat	-0.156*				-0.171*
	(0.091)				(0.094)
Low educated parents		-0.206*			-0.271*
		(0.117)			(0.143)
Low educated parents x Treat		0.280			0.286
		(0.185)			(0.221)
Low income parents			-0.133		-0.177
			(0.116)		(0.129)
Low income parents x Treat			-0.059		-0.071
			(0.185)		(0.203)
Single parent				0.048	-0.089
				(0.118)	(0.152)
Single parent x Treat				-0.147	-0.056
				(0.185)	(0.240)
Constant	-0.156***	-0.057	-0.085	-0.175**	0.100
	(0.057)	(0.074)	(0.081)	(0.077)	(0.161)
Observations	479	489	489	489	479
Adj. R-squared	0.057	0.021	0.020	0.015	0.063

Table A4.1: Interaction of treatment and parental background. The prosociality measures of mother and child are constructed using a principal component analysis for the aggregation of altruism, trust and other-regarding behavior, respectively. The prosociality measures are standardized, the other variables are dummies. Displayed coefficients are OLS estimates with robust standard errors in brackets. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Pre-treatment measures	Dependent Variable: Assigned to Treatment Low SES Group All Children		Follow-up Sample	
	(1)	(2)	(3)	(4)
Altruism	0.014		0.011	
	(0.020)		(0.021)	
Trust	0.008		0.002	
	(0.020)		(0.022)	
Other-regarding Behavior	-0.002		0.010	
	(0.020)		(0.022)	
Prosociality		0.010		0.016
		(0.020)		(0.022)
Observations	568	568	480	480
Pseudo R2	0.0008	0.0003	0.0009	0.0009
Wald test p =	0.898	0.625	0.913	0.475

Table A4.2: Check for baseline balance regarding target variables. Dependent variable is one if a child was selected into the Treatment Low SES group and zero otherwise. In column (1) and (2) all low SES children with valid measures in the pre-treatment interview are considered. The sample in column (3) and (4) is restricted to those how also took part in the post-treatment interviews. All measures are elicited before the families were informed about the assignment. The prosociality measure is constructed using a principal component analysis for the aggregation of altruism, trust and other-regarding behavior. To retain the structure of the measure over waves, the same weights are used to construct the pre-treatment measure as for the post-treatment measure. All independent variables are standardized. Displayed coefficients are average marginal effects of Probit estimates with robust standard errors in brackets. Wald tests retain the null hypothesis that all coefficients in the respective column are simultaneously equal to zero (p-values of chi2-tests are displayed); in line the Pseudo R2 never exceeds 0.001. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Pre-treatment measures	Dependent Variable: Lost to follow-up (1)	(2)
Treatment Dummy	-0.009	-0.006
	(0.031)	(0.032)
Altruism	0.007	
	(0.015)	
Altruism x Treatment Dummy	0.011	
	(0.026)	
Trust	0.023	
	(0.021)	
Trust x Treatment Dummy	0.020	
	(0.032)	
Other-regarding behavior	-0.008	
	(0.018)	
Other-regarding behavior x Treatment Dummy	-0.045	
	(0.034)	
Prosociality		0.004
		(0.019)
Prosociality x Treatment Dummy		-0.021
		(0.030)
Constant	0.158***	0.157***
	(0.019)	(0.019)
Observations	568	568
Adj. R2	0.0039	-0.0043
Wald test p	0.269	0.882

Table A4.3: Check for the absence of selective attrition. Dependent variable is one if a child is lost to follow-up, i.e. did not take part in the post-treatment interview, and zero otherwise. Lost to follow-up is regressed on pre-treatment measures of prosociality, assignment into treatment group and their interactions. The sample under study consists of all low SES children with valid measures in the pre-treatment interview. Prosociality variables are standardized and are elicited before the families were informed about the assignment. The prosociality measure is constructed using a principal component analysis for the aggregation of altruism, trust and other-regarding behavior. To retain the structure of the measure over waves, the same weights are used to construct the pre-treatment measure as for the post-treatment measure. Displayed coefficients are OLS estimates with robust standard errors in brackets . Wald tests retain the null hypothesis that all coefficients in the respective column (except for the constant) are simultaneously equal to zero (p-values of F-tests are displayed); in line the Adj. R2 never exceeds 0.004. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

	Prosociality	Altruism	Trust	Other-regarding behavior
	(1)	(2)	(3)	(4)
Treatment Dummy	0.299***	0.202**	0.235**	0.173*
	(0.092)	(0.095)	(0.098)	(0.095)
Dummy Cologne	0.017	0.065	0.105	-0.086
	(0.104)	(0.106)	(0.104)	(0.099)
Constant	-0.166	-0.145	-0.167*	-0.044
	(0.101)	(0.104)	(0.101)	(0.096)
Observations	489	492	494	491
Adj. R2	0.016	0.005	0.008	0.006

Table A4.4: Main analysis including city fixed effects (compare Fig. 4.2 and Fig. 4.3). City fixed effects are included since our experimental design used conditional random assignment (conditional on the place of residence of the families). The prosociality measure is constructed using a principal component analysis for the aggregation of altruism, trust and other-regarding behavior. All dependent variables are standardized. Displayed coefficients are OLS estimates with robust standard errors in brackets. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

	Prosociality	Altruism	Trust	Other-regarding behavior
	(1)	(2)	(3)	(4)
Treatment effect (TOT)	0.399***	0.251**	0.283**	0.263**
	(0.126)	(0.126)	(0.127)	(0.125)
Observations	489	492	494	491

Table A4.5: Treatment-on-the-treated (TOT) analysis using random group assignment as instrument for actual treatment. 133 of the 180 children we intended to treat were actually matched with a mentor. The main reason for not matching all ITT children was the lack of voluntary mentors. The prosociality measure is constructed using a principal component analysis for the aggregation of altruism, trust and other-regarding behavior. Coefficients are two-stage least-square (2SLS) estimates with robust standard errors in brackets. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

	Prosociality	Altruism	Trust	Other-regarding behavior
	(1)	(2)	(3)	(4)
Treatment Dummy	0.251***	0.173**	0.207**	0.169**
	(0.079)	(0.088)	(0.088)	(0.073)
Lagged Dependent Variable	0.515***	0.364***	0.359***	0.605***
	(0.039)	(0.040)	(0.046)	(0.043)
Constant	-0.137***	-0.085	-0.068	-0.113**
	(0.050)	(0.053)	(0.054)	(0.049)
Observations	475	480	494	489
Adj. R2	0.285	0.151	0.129	0.357

Table A4.6: Treatment effects conditional on baseline levels of outcome variables. The post-treatment measures are regressed on a treatment dummy and the pre-treatment measures. The prosociality measure is constructed using a principal component analysis for the aggregation of altruism, trust and other-regarding behavior. To retain the structure of the measure over waves, the same weights are used to construct the pre-treatment measure as for the post-treatment measure. All dependent variables and their lags are standardized by wave. Displayed coefficients are OLS estimates with robust standard errors in brackets. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Experimental protocols (translations from German)

Binary Dictator Game

”Have a look at these paper stars I’ve got here.“

→ Show the stars to the child.

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

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

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

“Here are two sheets of paper.“

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

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

→ Place the stars in the circles now.

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

→ Point to the respective circles.

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

→ Point to the first sheet.

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

→ Point to the second sheet.

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

Check the understanding of the rules

→ Repeat rules up to three times

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

“Okay, which sheet do you choose?“

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

“I will take care that the other child is getting something nice for the stars.”

Continuous Dictator Game A

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

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

"Look! Here are 6 stars."

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

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

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

→ Child puts the stars on the bags.

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

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

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

→ Remove the bags.

Continuous Dictator Game B

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

Again the bags are for the stars that you and the other child will receive."

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

"Look! Here are 6 stars."

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

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

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

→ Child puts the stars on the bags.

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

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

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

→ Remove the bags.

A5 Appendix to Chapter 5

Table A5.1: Relation between subjective health status and fairness perceptions (SOEP)

Dependent variable: subjective health status (higher values indicate better health)				
	(1)	(2)	(3)	(4)
Unfair wage	-0.180***	-0.169***	-0.199***	-0.262***
	[0.016]	[0.018]	[0.022]	[0.041]
Net wage /1000	0.054***	0.033***	0.033***	0.032*
	[0.006]	[0.008]	[0.012]	[0.018]
Age	-0.019***	-0.015***	-0.018***	-0.005
	[0.001]	[0.002]	[0.003]	[0.007]
Female	0.013	-0.041*	-0.050*	0.021
	[0.016]	[0.021]	[0.027]	[0.051]
Public sector		-0.042*	-0.013	0.097
		[0.025]	[0.033]	[0.062]
Tenure		0.000	0.000	-0.001
		[0.001]	[0.001]	[0.002]
Experience full time		-0.006***	-0.005	-0.006
		[0.002]	[0.003]	[0.005]
Experience part time		-0.003	-0.000	-0.001
		[0.003]	[0.005]	[0.007]
Realschule		0.029	-0.007	-0.014
		[0.024]	[0.031]	[0.057]
Fachoberschulreife		0.017	-0.022	-0.164*
		[0.038]	[0.050]	[0.087]
Abitur		0.059**	0.030	-0.069
		[0.030]	[0.041]	[0.073]
Other schooling degree		0.050	0.001	0.049
		[0.041]	[0.053]	[0.086]
No degree		-0.106	-0.169	0.427***
		[0.086]	[0.123]	[0.165]
In school		0.056	0.181***	
		[0.127]	[0.052]	
School info missing		0.028	0.018	0.033

	[0.052]	[0.075]	[0.121]
Lives in East Germany	0.030	0.078***	0.136***
	[0.020]	[0.026]	[0.050]
Self employed	0.069		
	[0.053]		
Firm size < 5	0.066*	0.054	0.051
	[0.034]	[0.054]	[0.109]
Firm size 6-10	0.017	0.015	-0.008
	[0.033]	[0.046]	[0.100]
Firm size 10-20	0.032	0.009	-0.062
	[0.035]	[0.045]	[0.092]
Firm size 101-200	0.022	0.062	0.102
	[0.033]	[0.040]	[0.071]
Firm size 201-2000	0.024	0.026	0.010
	[0.026]	[0.031]	[0.055]
Firm size above 2000	-0.005	0.017	0.028
	[0.027]	[0.032]	[0.058]
Firm size missing	-0.016	0.023	0.130
	[0.104]	[0.150]	[0.150]
Blue collar unskilled	-0.038	-0.024	0.053
	[0.052]	[0.085]	[0.133]
Blue collar craftsman	-0.003	-0.042	-0.022
	[0.037]	[0.043]	[0.082]
Blue collar foreman	-0.068	-0.095	-0.112
	[0.061]	[0.065]	[0.115]
Blue collar master	0.162	0.110	0.302
	[0.103]	[0.106]	[0.214]
White collar master	-0.119	-0.188	-0.232
	[0.123]	[0.130]	[0.186]
White collar skilled	0.016	-0.017	-0.010
	[0.041]	[0.056]	[0.110]
White collar unskilled	0.027	-0.056	-0.192
	[0.051]	[0.081]	[0.147]
White collar craftsman	0.093***	0.051	0.082
	[0.035]	[0.043]	[0.077]

White collar high qualified	0.141***	0.107**	0.176**
	[0.039]	[0.048]	[0.087]
White collar manager	0.046	0.038	0.180
	[0.067]	[0.079]	[0.135]
Civil servant low	0.385***	0.360**	-0.231*
	[0.142]	[0.159]	[0.123]
Civil servant intermediate	0.062	-0.028	-0.103
	[0.082]	[0.088]	[0.147]
Civil servant high	0.101*	0.091	0.119
	[0.061]	[0.073]	[0.116]
Civil servant excecutive	0.138**	0.091	0.139
	[0.069]	[0.084]	[0.126]
Other occupation	0.072*	-0.040	
	[0.040]	[0.140]	
Single	0.004	0.006	-0.034
	[0.023]	[0.029]	[0.092]
Widowed	0.032	0.010	-0.086
	[0.070]	[0.111]	[0.132]
Divorced	0.010	0.018	0.009
	[0.036]	[0.044]	[0.069]
Industry missing	-0.125	-0.066	0.005
	[0.081]	[0.116]	[0.177]
Industry energy	-0.139	-0.050	-0.015
	[0.088]	[0.108]	[0.184]
Industry mining	-0.234	-0.279**	-0.476***
	[0.145]	[0.132]	[0.172]
Industry manufacturing	-0.112*	-0.062	-0.139
	[0.064]	[0.088]	[0.137]
Industry construction	-0.107*	-0.015	-0.039
	[0.065]	[0.088]	[0.139]
Industry trade	-0.134**	-0.018	-0.102
	[0.065]	[0.090]	[0.150]
Industry transport	-0.189**	-0.094	-0.136
	[0.074]	[0.097]	[0.153]
Industry bank/insurance	-0.131*	-0.084	-0.280*

		[0.073]	[0.100]	[0.167]
Industry services		-0.114*	-0.041	-0.177
		[0.063]	[0.088]	[0.138]
Openness		0.016*	0.019	0.022
		[0.010]	[0.013]	[0.023]
Conscientiousness		0.064***	0.077***	0.036
		[0.010]	[0.012]	[0.024]
Extraversion		0.021**	0.016	0.022
		[0.009]	[0.012]	[0.022]
Agreeableness		0.049***	0.051***	0.113***
		[0.009]	[0.011]	[0.020]
Neuroticism		-0.103***	-0.102***	-0.115***
		[0.009]	[0.011]	[0.021]
	4.351***	4.334***	4.405***	3.803***
Constant	[0.030]	[0.091]	[0.126]	[0.343]
Observations	11,638	9,988	5,892	1,878
R-squared	0.080	0.120	0.132	0.100

OLS estimates with robust standard errors in brackets. The dependent variable measures subjective health status on a five-point scale from "bad" to "very good". ***, **, * indicate significance at the 1-, 5-, and 10-percent level, respectively. "Unfair wage" is a dummy variable equal to one if the respondent answered the question "Do you consider the income that you get at your current job as fair?" with "no" and zero otherwise. Additional controls include marital status (married (baseline category), single, widowed, divorced), whether the respondent lives in East Germany in 2009, labor market status (working in public sector, tenure, full time and part time experience), dummies for educational background (Hauptschule (baseline category), Realschule, Fachoberschulreife, Abitur, other schooling degree, no schooling degree, missing), dummies for firm size (self-employed, below 5, 6-10, 11-20, 21-100 (baseline category), 101-200, 201-2000, more than 2000, missing), occupational status (unskilled blue collar worker, skilled blue collar (baseline category), blue collar craftsman, blue collar foreman, blue collar master, white collar unskilled, white collar skilled, white collar craftsman, white collar master, white collar high qualified, white collar management, civil servant, civil servant intermediate, civil servant high, civil servant executive, other occupation), industry code (agriculture (baseline category), energy, mining, manufacturing, construction, trade, transport, bank/insurance, services, missing). Controls also include measures of personality (Big-5). The sample in column (1) contains all SOEP participants who are in any way active in the labor market in 2009. The sample in column (2) excludes individuals for whom not all controls are available or who just started in the current firm and whose work related information therefore does not refer to the current employer. The sample in column (3) is additionally restricted to dependent full-time employed individuals with positive income. In addition to the restrictions in column 3, the sample in column (4) is additionally restricted to individuals who are at least 50 years old. ***, **, * indicate significance of the "Unfair wage" coefficient at the 1-, 5-, and 10-percent level, respectively.

Panel data analysis

We complement the cross-sectional analysis and exploit the panel structure of the SOEP to develop dynamic panel data models which allow testing for a Granger causal effect of unfair pay on health outcomes. Causality in the sense of Granger (1969) implies that a potential effect from x on y is absent if lagged values of x_t add no further information to explain y_t beyond lagged values of y_t itself.[19]

The bivariate dynamic panel data model we use is adapted from Holtz-Eakin et al. (1988) and allows for individual fixed effects,

$$H_{it} = \sum_{l=1}^{h} \beta_l H_{it-l} + \sum_{j=1}^{k} \delta_j U_{it-j} + I_i + Y_t + u_{it} \tag{1}$$

where H_{it} is subjective health of individual i in period t ($i = 1, \ldots, N; t = 1, \ldots T$). H_{it} is explained by its own lags, the lags of the individual's perception of unfair pay (U_{it}), an individual fixed effect (I_i) and year dummies (Y_t); h denotes lag lengths of subjective health and k denotes lag lengths of fairness perception. The null hypothesis to be tested is that there exists no Granger causal effect of unfair wage perceptions on subjective health, i.e., that all δ_j are equal to zero.

The data structure of the SOEP does not allow constructing a dynamic panel data model for specific diseases because questions regarding specific diseases were only asked in 2009 and 2011. The question concerning subjective health status, however, was asked more often and we use data from 2001 to 2011. The survey question regarding perception of unfair pay (see section 5.3) was asked in the SOEP in the years 2009, 2007 and 2005. This data structure determines the period length to be two years which is conservative concerning the detection of Granger causality since causality may become effective faster than that, but using two-years-lags is common in health economics (see e.g., Michaud and Van Soest (2008)). Given this data structure, to maximize the number of estimable time periods and to hold the model as flexible as possible, we fix the number of lags of unfair pay perception (k) to one and calibrate the model by varying j, the number of lags of subjective health status. Using Arellano-Bover/Blundell-Bond estimators enables us to estimate the

[19] For an interpretation and discussion of Granger causality, see Hamilton (1994).

model for the years 2011, 2009 and 2007, with up to three lags of subjective health status.[20] In light of the medical literature (Roger et al., 2012) and our results in Table 5.3 (compare columns 3 and 4), we expect to observe the dynamic relation between health and unfair pay in particular for full-time employees who are older than 50 years. Therefore, we construct a balanced panel of individuals who work as full-time dependent employees in the years 2011, 2009 and 2007, and are born in 1961 or earlier. Thus, every individual in the sample is at least 50 years old in one or more periods. Estimation results are presented in Table A5.2. The estimation shown in column 1 includes one lag of subjective health; the estimations in columns 2 and 3 include two and three lags, respectively. The Hannan-Quinn information criterion Andrews and Lu (2001) selects the model with two lags of subjective health status (column 2) as preferred specification and the Sargan test of overidentification does not reject the validity of the instrumental variables in this specification ($p = 0.187$). A t-test rejects the null hypothesis that the coefficient of the lag of unfair wage perception is zero ($p = 0.025$) and therefore indicates a Granger causal effect of unfair wage perceptions on subjective health. This result is robust for reducing or increasing the lag lengths of subjective health (column 1 and 3) or extending the model by adding lags of net wages.[21]

[20]For validity of these estimators we have to assume that there is no serial correlation in the idiosyncratic errors . We cannot test for this assumption since the data structure limits our model to $T = 3$, and testing it requires $T \geq 5$ (Arellano and Bond, 1991).

[21]For example, adding one lag of net wage to the specification in column 2 of Table A5.2 does basically not change coefficients. While the lag of unfair wage is significant ($p = 0.021$), the lag of net wage is insignificant ($p = 0.943$). Results are available upon request.

Table A5.2: Dynamic panel estimation on the relation between perception of unfair pay and subjective health status

Dependent variable: subjective health status (higher values indicate better health)			
	(1)	(2)	(3)
Subjective Health$_{t-1}$	0.100***	0.170***	0.195***
	[0.021]	[0.038]	[0.052]
Subjective Health$_{t-2}$		0.072**	0.092**
		[0.032]	[0.044]
Subjective Health$_{t-3}$			0.022
			[0.030]
Unfair Wage$_{t-1}$	-0.089**	-0.092**	-0.094**
	[0.040]	[0.041]	[0.042]
Time Dummies	Yes	Yes	Yes
Sargan statistic	21.10**	14.91	14.07
Hannan-Quinn IC	-26.16	-28.41	-25.31
Number of Individuals	1,292	1,292	1,292

Arellano-Bover/Blundell-Bond linear dynamic panel estimations with standard errors in brackets. The balanced sample is restricted to dependent full-time employees who are born in 1961 or earlier. ***, **, * indicate significance at the 1-, 5-, and 10-percent level, respectively.

Instructions of the experiment

In the following we present a translation of the original German "employee" instructions.

Instructions for Employees

You are now taking part in an economic experiment. Please read the following instructions carefully. Everything that you need to know to participate in this experiment is explained below. Should you have any difficulties in understanding these instructions please notify us. We will answer your questions at your cubicle.

During the course of the experiment you can earn money. The amount of money that you earn during the experiment depends on your decisions and the decisions of another participant. At the end of the experiment you will receive the sum of money that you earned during the experiment in cash.

Please note that communication between participants is strictly prohibited during the experiment. In addition we would like to point out that you may only use the computer functions, which are required for the experiment. Communication between participants and unnecessary interference with computers will lead to exclusion from the experiment. In case you have any questions we are glad to assist you.

The participants of this experiment were randomly assigned the roles of employers and employees. You are an employee for the entire course of the experiment.

In the following you can earn money by working on a task. The money you earn will be received by your employer, who decides on how to divide the money between him and you. The interaction is completely anonymous, i.e., at no point you will learn the identity of the employer and the employer will not learn your identity.

Your work task

The work task is to count the correct number of zeros on prepared sheets containing zeros and ones. At your cubicle you find an example of such a sheet. At the top you see the sheet number. Below that you find a table with zeros and ones. To earn money, you have to count the correct number of zeros and enter it into the computer. To do that you will receive a new computer screen for each sheet.

The first input screen is for the first sheet. Under the heading: "How many zeros are on sheet 1?" you find a box where you can enter a number. Type the correct number into that box and click on "OK". As soon as you have clicked on the "OK"-button, the screen for the next sheet appears etc.

As long as you have not clicked on the OK-button, you can change your entry. As soon as you have clicked on OK, however, the next screen appears.

For each correctly solved sheet you create revenue of 3 Euro. For example, if there are 29 zeros on a particular sheet and you type 29, you create revenue of 3 Euro. If your entry deviates by plus/minus 1 from the correct number of zeros, you receive 1 Euro. If your entry deviates by more than plus/minus 1, you create no revenue for that particular sheet.

Example:
Suppose, the correct number of zeros on a particular sheet is 15.
If you type 15, you create revenue of 3 Euro.
If you type in either 14 or 16, you create revenue of 1 Euro.
If you type in a number smaller than 14 or larger than 16, you create revenue of zero Euro.

Please note: As soon you have clicked OK, you cannot revise your entry anymore. The next screen for the next sheet appears immediately.

On each input screen you are informed about the number of correctly solved sheets, the number of almost correctly solved sheets (deviation plus/minus 1) as well as the resulting amount of revenue you have produced. In addition you see on the screen the remaining time in seconds.

You have 25 minutes to solve sheets and create revenue (25 minutes = 1500 seconds).

You can work on as many sheets as you like: None, one, two etc. up to a maximum of 20. The sheets will be allocated as soon as you have read the instructions.

The decision of the employer

Your employer will receive the amount of money you have produced. He divides the amount of money between himself and you. Any feasible allocation is possible. For example, the employer can keep the whole amount for himself, give the whole amount to you, he can keep 10 percent of the amount and give you 90 percent, he can divide exactly equally etc.

The employer does not work and does not create any revenue. He knows, however, that the amount of money that he can divide depends on your work effort.

Following your working time and the allocation decision of the employer, you will have to complete a short questionnaire. Then, the experiment is over and you will receive your payments in cash, depending on the amount of money and the allocation decision. If you have any questions, please let us know.

If you have read these instructions, please click "Start".