

A Nation-Wide Laboratory

Examining trust and trustworthiness by integrating behavioral experiments into representative surveys

By Ernst Fehr, Urs Fischbacher, Bernhard von Rosenblatt,
Jürgen Schupp and Gert G. Wagner*

Abstract

Typically, laboratory experiments suffer from homogeneous subject pools and self-selection biases. The usefulness of survey data is limited by measurement error and by the questionability of their behavioral relevance. Here we present a method integrating interactive experiments and representative surveys thereby overcoming crucial weaknesses of both approaches. One of the major advantages of our approach is that it allows for the integration of experiments, which require interaction among the participants, with a survey of non-interacting respondents in a smooth and inexpensive way. We illustrate the power of our approach with the analysis of trust and trustworthiness in Germany by combining representative survey data with representative behavioral data from a social dilemma experiment. We identify which survey questions intended to elicit people's trust correlate well with behaviorally exhibited trust in the experiment. People above the age of 65, highly skilled workers and people living in bigger households exhibit less trusting behavior. Foreign citizens, Catholics and people favoring the Social Democratic Party or the Christian Democratic Party exhibit more trust. People above the age of 65 and those in good health behave more trustworthy or more altruistically, respectively. People below the age of 35, the unemployed and people who say they are in favor of none of the political parties behave less trustworthy or less altruistically, respectively.

Zusammenfassung

Die experimentelle Ökonomie führt typischerweise Labor-Untersuchungen durch, die mit homogenen und selektiven Versuchspersonen arbeiten. Repräsentative Surveys leiden hingegen unter Messfehlern und der Frage, ob hypothetisches Verhalten, das erhoben wird, mit tatsächlichem Verhalten korrespondiert. Deswegen präsentie-

* Corresponding authors: Ernst Fehr and Gert G. Wagner. This paper would not have been written without the opportunities for collaboration provided by the Institute for Advanced Study Berlin (Wissenschaftskolleg zu Berlin). The idea for this project was born when Ernst Fehr was a Fellow at the Wissenschaftskolleg during the academic year 2001 / 2002. The specific method for combining surveys with behavioural experiments proposed here originated in discussions between Richard Hauser (also a fellow at the Wissenschaftskolleg at this time), Gert G. Wagner and Ernst Fehr. We are grateful to Andreas Stocker who works for Infratest Sozialforschung and to two referees. The usual caveats apply.

ren wir eine Methode, mit der man die Schwächen beider Ansätze überwindet, indem man Survey-Daten mit Experimenten kombiniert. Die von uns entwickelte Methode hat den Vorzug, dass sie es auf preiswerte Art und Weise erlaubt, Experimente, die die Interaktion zweier Personen erfordern, mit Standard-Surveys zu kombinieren, die auf der Unabhängigkeit der befragten Personen beruhen. Wir illustrieren die Möglichkeiten unserer Methodik mit einer Untersuchung über Vertrauen und Vertrauenswürdigkeit in Deutschland, bei der eine repräsentative Umfrage mit einem Experiment über eine Dilemma-Situation verbunden wurde. Unsere Resultate zeigen zum einen, dass nicht alle Survey-Fragen über Vertrauen mit den Ergebnissen des Experiments korrelieren. Zum zweiten zeigen unsere Ergebnisse, dass ältere Menschen (65 Jahre und älter), Hochqualifizierte und Personen, die in größeren Haushalten leben, im Experiment unterdurchschnittlich Vertrauen offenbaren. Für Ausländer, Katholiken und die Anhänger der großen Volksparteien finden wir überdurchschnittliches Vertrauen. Vertrauenswürdigkeit bzw. Altruismus zeigen im Experiment Ältere und Befragte mit gutem Gesundheitszustand. Unterdurchschnittliche Vertrauenswürdigkeit bzw. Altruismus zeigen jüngere Befragte (bis zu 35 Jahren), arbeitslos Gemeldete und Befragte ohne eine Parteineigung.

JEL Classification: A 13, C 42, C 82, C 92, C 93, D 84, J 24

1. Advantages and Disadvantages of Laboratory Experiments

For a long time the social sciences were viewed as a field that is based on theoretical reasoning and passive observation of empirical facts. Social scientists thought that it is impossible to use controlled laboratory experiments to enhance our understanding of human behavior, economic mechanisms, social institutions and government policies. Laboratory experiments first became firmly established in social psychology whereas in economics, sociology and political science they remained somewhat marginal until today. In economics views gradually changed during the last 3 decades – first very slowly and more recently change has been rather rapid. The award of the Nobel Prize in Economics to Vernon Smith, who is probably the most influential pioneer in Experimental Economics, is an indication of the fact that laboratory experiments are now routinely used by an increasing number of economists.

The key advantages of laboratory experiments are the tight control of the environment under which subjects make their decisions and the replicability of the data. If somebody does not believe (in the robustness of) the behavior pattern observed in an experiment he or she can just replicate the conditions under which the original data have been generated, and observe whether the original data pattern does or does not emerge. If somebody does not believe the proposed interpretation of an observed data pattern he or she can design new experiments to test the original interpretation against competing interpretations. If one is interested in the causal impact of certain variables then the experimenter can vary the variable of interest and

observe how this variation affects subjects' behavior. In fact, the exogenous variation of variables in controlled environments is the only truly reliable way to make causal inferences. If somebody believes that an important factor has been left unspecified or uncontrolled, or that this factor could not play a role in the experiment although in the external world it is likely to play a role, it is often possible to change the experimental conditions in such a manner that the factor that had initially been left out can now play a role.

Thus, there can be little doubt that laboratory experiments provide the chance to substantially enhance our knowledge about human behavior. It is also clear, however, that laboratory experiments have their limits. One of the most frequent criticisms of experiments is that they are „artificial“ and that they lack „external validity“. However, critiques do not always indicate with sufficient clarity what is meant by these criticisms. Sometimes the simplicity of experiments is attacked because the real world – which we ultimately want to understand better – is so much more complex. This criticism is often misplaced because, in general, it is necessary to understand the simple cases first before one is able to understand the more complex cases. As good theory starts from understanding simple cases good experiments also take simple cases as their starting point. For example, much of the progress in modern genetics arose from the analysis of very simple organisms with an extremely limited number of genes.

Experimental results are externally valid if there is good reason to believe that the experimental environment, under which the results have been generated, captures essential elements of naturally occurring environments. Again, the similarity with good theory is illuminating here because theories that are thought to be „relevant“ must capture essential elements of the external situation to which they are applied.

More important criticisms of experiments concern the constrained subject pool that is used in most laboratory experiment and the selectivity of the subject pool. Most laboratory experiments are run with students as subjects. It is obvious that students' behavior may differ from the behavior of other groups in society. Students probably come from families with above average income, are of a certain limited age range, have above average IQs and are more used to abstract problem solving. Moreover, it is even unclear whether the typical student subject pool used is representative of the student population.

There are a few studies that examine the potential biases caused by using students as laboratory subjects. Cooper, Kagel, Lo and Gu (1999), for instance, compare the behavior of Chinese middle managers and Chinese students in an experiment examining the ratchet effect. The ratchet effect arises if managers of firms, which are regulated by central planning authori-

ties, withhold effort because they fear that if they perform too well, the authorities will ratchet up the planning targets for next year. Cooper et al. show that over time students and managers converge to the same behavior but initially there are some behavioral differences across these groups.

While studies like the one by Cooper et al. are very interesting and useful they do not fully address all the potential biases created by using student subjects. Nor do they address the self-selection problem inherent in the typical recruiting procedures for laboratory experiments. Only fully representative experiments can address these problems.

This paper will present a method for overcoming these problems by combining an experiment with a nationwide survey. Doing so we discuss the advantages and disadvantages of surveys in section 2. In section 3 we present strategies for integrating experiments into surveys. In other words: we present the methodology for a nation-wide laboratory and we apply this method by examining trust and trustworthiness for a representative sample of the population in Germany. In section 4 we describe this first integration of a social dilemma experiment that requires interaction between respondents in a survey of non-interacting respondents. In section 5 we discuss preliminary empirical results and section 6 concludes.

2. Advantages and Disadvantages of Surveys

The main aim of surveys is to draw conclusions for a universe of a population (in the social sciences either for persons, households or firms) without surveying the entire population. The samples should be randomly chosen because otherwise representativeness cannot be ensured.

Survey data can be collected by interviewing respondents or by the analysis of register data. Register data are very precise if the data provide the basis for administrative purposes. For example, the files with the contributions paid to the statutory pension insurance are a good indicator of the insured persons' earnings. The major shortcoming of register data is the complete lack of subjective data. Survey data, which is collected by means of interviews¹ is not as reliable as register data with regard to objective infor-

¹ There are several interview methods at hand. The traditional method is a face-to-face interview by „paper and pencil“ (PAPI). If an interviewer fills in the questionnaire in a laptop the method is called „Computer Assisted Personal Interview“ (CAPI). If the interview takes place through the phone in most studies the answers are typed in a computer system (Computer Assisted Telephone Interview – CATI). Until recently, self-completed surveys have been administered by ordinary mail. Internet surveys are a modern form of mail surveys. In Internet surveys the respondent fills in a questionnaire via the Internet. However, in principle Internet surveys are just a new kind of self-completed surveys. To the extent to which randomly selected participants

mation, like earnings, but the interview technique allows to ask questions about the subjective well-being of respondents and their beliefs and expectations.

Thus, the main advantage of interview-based survey data is that one can ask people a host of questions about their objective situation, about their subjective feelings, their satisfaction with life and work, their health status, their behavioral intentions, and the like. The main disadvantage of surveys is the potential imprecision of many data, the uncontrolled manner in which the data are gathered and – compared to experimental settings – the lack of economic incentives. For instance, saying that I am willing to pay X Euros to save an endangered species is one thing, actually paying those X Euros may well be another thing. Answers to sensible questions about people's values may be biased due to social desirability bias. In addition, answers to questions about past behavior may be biased due to limited memory.² Words can mean very different things for different people and the behavioral content of people's answers is often not clear.

If one is, for instance, interested in the amount of trust in other people's altruism, honesty and fairness one may ask questions like the following:

Do you think that most people

- *try to take advantage of you if they got a chance*
- *or would they try to be fair?*

Or one can ask people to what extent they agree or disagree with the following statements.

- a) In general, one can trust people*
- b) In these days you can't rely on anybody else*
- c) When dealing with strangers it is better to be careful before you trust them.*

It is unclear to what extent the answers to such questions are good indicators for people's real trusting behavior. Yet, these questions are mainly interesting because they try to capture something about people's real behavior. Answers to attitudinal trust questions have been used in cross-country regression of the impact of social capital on economic growth (Knack and Keefer 1997). In view of the uncertain behavioral validity of the answers to

are unwilling to participate in a mail survey self-selection problems emerge. In most Internet surveys the self-selection problem is even more severe because even the initial sample of potential respondents is not randomly chosen. The respondents of Internet surveys select themselves into the „sample“.

² Research on nominal wage rigidity indicates that survey based income measures are plagued by considerable measurement error (see Altonji and Devereux 1999, Fehr and Goette 2000). However, this problem can be solved by collecting panel data, i.e. surveying the same unit over time.

these questions it requires a big leap of faith to believe in the credibility of these regressions.

In a recent study Glaeser et al. (2000) showed for a sample of students that attitudinal trust questions like the ones above do *not* predict real trusting behavior in controlled experiments. The authors find that they predict people's trustworthiness but not people's trusting behavior. However, the study by Glaeser et al. also nicely illustrates one of the important limits of most previously conducted experiments. The sample population for this study consisted of Harvard Undergraduates. It seems natural, therefore, to question the generality of these results. Perhaps, attitudinal trust questions tell us more about people's real trusting behavior if one uses a representative sample.

Yet, despite the use of a rather limited subject pool the Glaeser et al. study shows some of the major advantages from combining survey studies with experiments. The behavioural content of survey questions can be validated and the socio-economic and demographic determinants of experimental behavior can be examined.

3. Combining Surveys with Experiments

In this section we propose a method allowing a smooth and inexpensive integration of simple game-theoretic experiments into a survey design. After some general considerations we will discuss two basic methods: the „strategy method“ and the „specific decision method“. Then we show how we implemented the “specific decision method” into our survey.

A survey typically consists of a large number of randomly chosen independent respondents. There is no interaction between the survey respondents. Thus, it is relatively easy to combine survey responses with experimental responses if the participants of the experiment do not interact with each other. Individual decision-making experiments like, for instance, experiments that elicit individual risk aversion measures or individual discount rates can, therefore, be easily added to a survey interview. Nevertheless this method is infrequently used.³ The situation is significantly more complex, however, when pair-wise experimental interactions between two survey participants like, for instance, in a Prisoners' Dilemma (PD) take place.

In the case where interaction between pairs involves simultaneous – and independent – decisions the situation is still relatively easy to handle. The pairs of players can be ex-ante matched. Then the people who conduct the

³ We do not know why this is the case. We believe the reason is a de-facto „separation“ of the research communities who are either doing survey or experimental research in the social sciences.

interviews can, after they have carefully explained the experiment to the survey participant, collect the experimental decisions made by individual survey participants. In the case of a PD the interviewer would thus collect the information whether the survey participant decided to cooperate or to defect. After the experimental decisions have been collected the survey participants can be informed by mail about the decision of their counterpart. Payment can also be made by ordinary mail.

The situation is even more complex if pair-wise interactions take place sequentially, i.e., if there is a first-moving player who makes a decision and this decision is communicated to a second-moving player, who then makes his decision, and both players do not know each other. Many of the most interesting interactions have a sequential structure so that it is important to combine sequentially played experiments with survey methods. In the following we describe a method how this can be achieved. To facilitate understanding we do this in the context of a sequentially played social dilemma game. In this game player A and player B are matched. Both are endowed with 10 €. Player A can send any amount between 1 and 10 € to player B. The experimenter doubles the amount sent, that is, if A sends x €, player B receives $2x$ €. After player B has been informed about the amount sent by A, B can send any amount between 1 and 10 €, which is then also doubled by the experimenter. Thus, if B sends y € to A player A receives $2y$ €. The total payoff for player A is thus given by $10 - x + 2y$ €; the total payoff for B is given by $10 + 2x - y$ €.

In the experimental literature the gift exchange game (Fehr, Kirchsteiger and Riedl 1993) or the trust game (Berg, Dickhaut and McCabe 1995) have such a sequential structure. These games mimic the effort enforcement problem in labor markets or the quality enforcement problem in goods markets and the repayment enforcement problem in credit markets. Thus, some of the most important economic interactions have the structure of a social dilemma. Our game described above can also be interpreted as a simple sequential trade under incomplete contract enforcement. For A the goods (€) that are owned by B are worth twice as much as for B, and vice versa. Both players would be better off if they trade their goods but player A has to have trust in B to be willing to trade and B has to be trustworthy.

One way to implement our sequential social dilemma game in the context of a survey would be to first interview one half of the respondents – the first movers (players A). Then the other half of the respondents – the second movers (players B) – is interviewed. Each second mover is informed about the decision of „his“ of „her“ first-mover after which the second-mover makes the decision. This decision of player B is then communicated to the matched player A. However, this sequential way of implementing the game imposes

heavy administrative and logistical problems on the organization of the survey. It is, therefore, desirable to free the implementation of the survey from this burden and conduct the sequential game in a simultaneous way, i.e. a one step fieldwork procedure.⁴ This can, in principle, be achieved in two ways, the „strategy method“ and the „specific decision method“.

In our sequential social dilemma the strategy method requires that player B makes a decision for each possible transfer of player A. Player B makes this decision before he knows the actual transfer that has been chosen by player A. Ex-post, when both players have made their decision player B is informed about the actual transfer made by player A and player A is informed about the actual transfer of player B. The actual transfer of B is given by the amount sent by B in response to the actual transfer of A.

The main advantage of the strategy method is that it provides more information about the behavior of player B than the „specific decision method“ which is described below. If one combines survey methods with experimental methods the strategy method gains an additional advantage by rendering it easier to achieve this combination for sequential games. However, the strategy method may also have some disadvantages. One potential disadvantage is the dilution of incentives. In the case of our particular sequential social dilemma the probability that any one of the 11 feasible actions is chosen by A is well below 1.⁵ Another potential disadvantage is that asking B how he would respond in any of the 11 feasible situations is less emotionally arousing than if B is confronted with an actual decision by player A. It may be one thing to actually get a generous transfer from player A but another thing to be asked how one would respond to a generous transfer by A.⁶ Psychological consistency requirements might also affect B's response to A under the strategy method. For instance, the need for consistency might induce player B to prefer a monotonically increasing transfer – giving more the more A gives – whereas under the „specific decision method“ the need for consistency cannot become operative. Under the specific decision method B is confronted with the actual decision of A and responds only to this decision.

So far the experimental literature does not provide clear-cut answers about whether the strategy method induces sizeable and significant distortions relative to the more natural specific decision method. All the same, if one has a large sample of first and second movers the need for the strategy

⁴ We tested both methods by means of a pretest in April 2002.

⁵ Yet, research on the effects of stake size on experimental outcomes reveals that this effect is likely to be small and not too important (Camerer and Hogarth 1999).

⁶ The gap between hot and cold cognitions might be especially important in bargaining games in which negative reciprocity applies (like in the ultimatum game) since negative emotions causing punishment might be stronger than positive emotions causing reward.

method is weakened because the likelihood that each of the feasible first-mover actions is chosen with sufficient frequency is relatively high. This raises the question whether it is possible to implement a sequential game within a survey in a simultaneous manner without the use of the strategy method. In the following we will present a solution to this problem.

Our method rests on the idea that the experimenters determine – based on previous experimental knowledge about the distribution of first mover actions – an ex-ante distribution of first mover actions. Every player B is confronted with a randomly chosen first mover action from an ex-ante distribution and he only responds to this randomly chosen action. For the decisions of players B and the final payoffs it does not play a role that the first-mover action has been randomly chosen from an ex-ante distribution. It is sufficient that every player B knows that he is matched with a randomly chosen other (anonymous) participant of the survey who has been assigned the role of player A. This method allows the administrators of the survey to simultaneously conduct the experiments with players A and B in a one-step procedure and to match players A and B ex-post, i.e., after all As and Bs have made their decisions.

This matching is done as follows. First, the actual distribution of first-mover choices is determined. For any feasible action x of the first movers there is a number of first movers who actually chose x and a number of second movers who responded to x . If the ex-ante distribution is based on sound information about previous first-mover behavior⁷, and if the sample size is large enough, the number of first- and second movers will be roughly equal for any x . If the numbers are exactly equal for a given x each second mover, who faced x , will be randomly assigned to one of the first movers who chose x . If the number of second movers (for a given x) is larger than the number of first movers, each first-mover is randomly matched with one of the second movers. In this case there will be a few second-movers who will not be matched with a first mover (however, the payoff for those second movers is clear, because they respond to a given x).⁸

In the opposite case, when, the number of first-movers (for a given x) is larger than the number of second movers, some of the second movers will be

⁷ We conducted a pilot survey in early 2002, which provided the basis for our ex-ante distribution of first-mover actions. The question of whether it would be better to generate the ex-ante distribution by an experiment did not apply because the pilot survey delivered a distribution which looked very much like distributions which are based on laboratory experiments.

⁸ One can also think of this situation as one in which some of the first-movers are matched with more than one second-mover. However, each first-mover gets paid only according to *one* randomly assigned second-mover. Thus, while a first-mover's decision may be decisive for more than one second mover, each first-mover is only paid according to the decision of *one* randomly assigned second mover.

repeatedly „matched“ with a first-mover. This means that the decisions of some individual second movers become decisive for more than one first-mover. However, each second-mover is paid according to only *one* randomly assigned first-mover.

The matching procedure can formally be described as follows: Let $M1_x$ be the number of first movers who choose x and $M2_x$ the number of second movers who face x . The first movers who have chosen x are randomly assigned the numbers between 1 and $M1_x$; the second movers who faced x get randomly assigned the numbers between 1 and $M2_x$. For a first mover with assigned number $m1_x$, the transfer of the second mover with assigned number $m2_x$ is decisive if

$$(1) \quad m2_x = 1 + (m1_x - 1) \bmod M2_x .$$

The advantage of the ex-post matching process is that one can combine the survey with a sequential experiment without any distortion of the usual way of conducting surveys. Moreover, the experimenter tells the subjects the truth about the game. Each subject is in fact matched with one randomly assigned participant of the survey who is in the other role and the decision of the counterpart, with which one is confronted, is in fact made by a real individual. The only information that the second movers do not have is that the action with which they are confronted may not yet have been made. As the survey-experiment is actually being conducted it may sometimes happen that the (ex-post) matched first-mover has already made his or her decision whereas at other times this decision will be made later. However, from a game-theoretic viewpoint this design feature is completely innocuous. What counts is that the second mover will be matched with a real person who has made a real decision and that the action of the second-mover does affect his own and the matched person's income.

4. First Implementation

The first implementation of our method of combining a survey with an experiment is based on a representative sample of adult individuals living in private households within Germany. The questionnaire is basically the one of the German Socio-Economic Panel Study (SOEP)⁹, being extended by several questions on trusting behavior. Our survey is a random sample based on the random walk method (cf. Fowler 1988).¹⁰ The survey-experiment

⁹ See Schupp and Wagner (2002) who also mention the possibility of surveying via the Internet. This method would give straightforward possibilities for combining a survey with experiments. But the representativeness of internet surveys is still a problem.

¹⁰ Details of this method are described in the appendices.

took place during June 28th – July 20th in 2002. 442 individuals were interviewed and 429 of them were willing to participate in the experiment.

In this study we found it useful to first conduct the full survey questionnaire. When all questions had been answered the interviewer explained the experiment with the help of written instructions to the survey participant.¹¹ No survey respondent was forced to participate in the experiment but 97% of the survey participants were willing to do so. After the interviewer had explained the rules of the experiment the participant received a decision sheet and an envelope.

The first-movers indicated how much money they transfer to the second mover. *After* they had made their decision they also indicated their *expectation* about the second mover's transfer. On the decision sheet of the second mover a hand written number indicated the transfer of the first-mover. The second-mover indicated on his sheet how much money he transfers to the first-mover. After a player made his decision he folded his decision sheet, put it into the envelope and closed the envelope. This means that the interviewer did not know the decision of the participant. When the interviewers explained the rules of the experiment they also told this feature to the participant, i.e., the participants knew that the interviewers will not know their decisions. The interviewers were also instructed to let the participants privately make their decisions.

Finally, in our case we also had the advantage that a well-established research institute – Infratest Sozialforschung – conducted the survey and the experiment. Infratest Sozialforschung is well known in Germany – it is one of the major institutes conducting opinion polls before elections and it routinely conducts many surveys each year. Infratest, in particular, administers every year the questionnaire for the German Socio-Economic Panel Study (SOEP) and has a reputation of being a reliable and trustworthy research institute. This is important because it enhances the credibility of the experimental instructions.

When all surveys and interviews had been conducted Infratest matched first- and second-mover, computed individual earnings and sent checks by snail mail to the participants. The first-movers were also informed about the decision of the matched second-mover.

In the following we present some of the results that emerged from a preliminary analysis of the data. First we shortly discuss to what extent our sample is representative for the German population. Then we examine whether

¹¹ The questionnaire of the survey can be downloaded at: http://www.diw.de/deutsch/sop/service/fragen/personen_pretest_2003.pdf. The instructions for the experiment can be viewed in appendix 2 of the DIW Discussion Paper No. 319 at: <http://www.diw.de/deutsch/publikationen/diskussionspapiere/jahrgang02/>.

our survey questions designed to elicit the participants' trust correlate with our behavioral measure of trust – the first movers' transfers. Then we analyze socio-economic and demographic determinants of trust and trustworthiness as exhibited in the experiment.

5. Empirical Results

In this section we address two questions: (1) To what extent do survey-reports of trusting behavior correlate with experimental trust behavior? (2) Who are the trusting and trustworthy people measured by socio-economic and demographic variables? Before we do so Table 1 shows the distribution of important demographic characteristics of the realized sample and compares it with the overall distribution of these characteristics in the overall sample of the SOEP, which comprises about 12,000 households and 25,000 individuals. The results of this comparison indicate that our sample is representative in all dimensions displayed in Table 1. The small differences between our sample and the SOEP sample are insignificant.

5.1 Do answers to trust questions predict trusting behavior?

In our survey we had several trust and fairness questions. In total there were 134 questions in the survey and our trust and fairness questions were answered at the beginning (Questions 5–10, see footnote 11). Our questions were taken from the General Social Survey conducted in the US, from Glaeser et al. (2000) and from other surveys.¹² In the following we describe them in some detail:

Question 5: Do you think that most people

- *try to take advantage of you if they got a chance*
- *or would they try to be fair?*

Question 6: Would you say that most of the time

- *people try be helpful*
- *or that they are mostly just looking out for themselves?*

Question 7: a) In general, one can trust people

b) In these days you can't rely on anybody else

c) When dealing with strangers it is better to be careful before you trust them.

¹² We are grateful to Wolfgang Jagodzinski for providing us with information about those surveys.

Table 1
Representativeness of sample

	Total sample Summer 2002 N	Distribu- tion Summer 2002 (in %)	Distribu- tion SOEP 2001 (in %)
<i>Gender</i>			
Women	241	0.55	0.51
Men	200	0.45	0.49
<i>Age</i>			
16 – 19	23	0.05	0.02
20 – 29	57	0.13	0.10
30 – 39	81	0.18	0.22
40 – 49	69	0.16	0.20
50 – 59	60	0.14	0.16
60 – 69	73	0.17	0.17
70 – 79	63	0.14	0.10
80 +	16	0.03	0.03
<i>Household size</i>			
1 person	125	0.29	0.26
2 persons	154	0.35	0.35
3 persons	68	0.16	0.17
4 persons	65	0.15	0.16
5 and more persons	26	0.06	0.07
<i>Regional distribution</i>			
big cities with 500,000 more	193	0.44	0.39
big cities with 100,000 to 500,000	73	0.17	0.18
cities with 50,000 to 100,000	23	0.05	0.05
countrified areas with 5,000 to 50,000	103	0.23	0.24
rural areas- less than 5,000	50	0.11	0.15

Question 7 had four answer categories: Agree fully, agree somewhat, disagree somewhat, fully disagree.

Question 8: In the following you are asked to which persons, groups and institutions you have more or less trust.

Question 8 contained 14 items, ranging from trust to your own family, to neighbors, friends, colleagues at the workplace, to schools, the police, the courts and big enterprises. The answer categories were: very much trust, relatively much trust, little trust, no trust at all.

Question 9: Have you ever spontaneously benefited from a person you did not know before?

Question 10: How often does it happen

a) that you lend personal possessions to your friends (CDs, books, your car, bicycle etc.)?

b) that you lend money to your friends?

c) that you leave your door unlocked?

Via factor analysis we built five factors out of Questions 5, 6, 7, 8 and 10. The higher a factor the more people exhibited trusting behavior in the past (Question 10), the more they express generalized trust in their family, their neighborhood, the police, the courts, etc. (Question 8), the more they believe in other people's fairness and altruism (Question 5, 6 and 7), and the more they have spontaneously benefited from other people's generosity (Question 9).

In Table 2 we present ordered probit regressions of first-mover and second-mover behavior on these trust factors.¹³ Regression (1) shows that Questions 7 and 10 have a significant impact on first-mover's transfers whereas the answers to the other questions do not correlate significantly with real behavior. This suggests that questions about past trusting behavior (Question 10) and about trust in strangers (Question 7) do predict trusting behavior in our experiment. Because we asked the first-movers in our experiment about the transfer they expected from the second-movers this expectation variable can itself be interpreted as a measure of trust. It is, therefore, interesting to examine whether, when controlling for the expected transfer, our trust questions still remain significant (see model (2) in Table 2). The results for model (2) indeed show that Question 7 becomes now insignificant. This suggests that trust in strangers, as exhibited in Question 7, becomes operative via more optimistic expectations about the second-movers transfer. However, Question 10 remains highly significant even after controlling for expectations: For given expectations about second movers' transfers those individuals who reported more trusting behavior in the past also exhibited more trust in our experiment. This fact may indicate that people who score high on question 10 are habitually more trusting or that they are more altruistic. For the same probability of getting things back they may be more willing to lend money and other personal possessions to their friends.

In order to assess the relevance of different questions as a predictor of trusting behavior it is also useful to examine the simple raw correlations between respondents' answers and respondents' behavior. This is done in

¹³ Due to relatively high non response to answers on questions about trust in other persons (factor 8) the numbers of cases are much smaller than in Table 4.

Table 2
The behavioral relevance of questions about trust

	1 st movers transfer (1)	1 st movers transfer (2)	2 nd movers transfer (3)
factor for questions 5 & 6: belief about people's fairness	0.251 (0.177)	0.053 (0.182)	0.192 (0.169)
factor for questions 7: belief in trustworthiness of others	0.325 (0.165)**	0.256 (0.160)	-0.112 (0.170)
factor for questions 8: trust in others and institutions	-0.161 (0.099)	-0.181 (0.103)*	0.008 (0.112)
question about whether person benefited from generosity of stranger in the past	-0.014 (0.203)	0.240 (0.211)	-0.296 (0.246)
factor for questions 10: frequency of past trustful behavior (e.g., lending to friends)	0.408 (0.121)***	0.361 (0.124)***	0.093 (0.122)
belief about the transfer of the 2 nd mover received from 1 st mover		0.261 (0.055)***	0.090 (0.031)***
Observations	147	137	141

Note: Ordered probit regressions. Robust standard errors in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3, which shows the Spearman Rank Correlations between answers and behavior. The table confirms that questions about past trusting behavior are by far the best predictor of trusting behavior in the experiment. One can also see that the correlation regarding answers to questions 7 (belief in others' trustworthiness) and to question 9 (having benefited from a stranger in the past) are significantly positively. Only the factor for the items of question 8 does not co-vary significantly with trusting behavior. However, this could also be due to the fact that the factor related to the different items of question 8 confuses trust in family members, neighbors and colleagues at the job with trust in institutions. Perhaps, if one separates these two components of trust, part of the answers to questions 8 is significantly related to trusting behavior. Yet, as Table 3 shows, separate factors for trust in persons and trust in institutions are also not significantly related to trusting behavior in the experiment.

In model (3) we examine the impact of our trust variables on second-movers' transfers. If one takes our questions literally, there is no reason to expect that there is a correlation between the answers on trust and second-movers' behavior which shall be based on trustworthiness respectively fairness. However, Glaeser et al. (2000) recently found that the answers to attitudinal questions about trust such as, for instance, to Question 7 was, are

Table 3
**Spearman Rank Correlations between trust questions
 and first-mover transfers**

	Spear- man's rho	<i>p</i> value	Number of observa- tions
factor for questions 5 & 6: belief that people are fair	0.1137	0.0987	212
factor for questions 7: belief in trustworthiness of others	0.1391	0.0435	211
factor for questions 8: trust in others and institutions	-0.0525	0.5232	150
Factor for first 4 items of question 8: trust in other persons	0.0233	0.7670	164
Factor for last 10 items of question 8: trust in institutions	0.0150	0.6360	192
question about whether person benefited from generosity of stranger in the past	0.1466	0.0324	213
factor for questions 10: frequency of past trustful behavior (e.g., lending to friends)	0.2898	0.0000	213

not correlated with trust but with trustworthy behavior in his experiment. In our representative sample this is, not the case. None of the trust questions comes close to be a significant predictor of trustworthy behavior.

Thus, taken together we have identified two survey measures of trust that correlate with trusting behavior in the experiment: direct questions about trust in strangers and questions about past trusting behavior.¹⁴ None of the survey measures of trust are good predictors of trustworthiness in the experiment.

5.2 Socio-economic and demographic determinants of trust and trustworthiness

Our sample is representative of the German population and we know a lot about the socio-economic and demographic characteristics of our respondents. In the following we illustrate how this information can be profitably used to learn more about the determinants of trust and trustworthiness. We would like to stress, however, that the analysis below is rather preliminary. It does not constitute a fully-fledged examination but is intended to illustrate the usefulness of our approach.

¹⁴ Glaeser et al (2000) also found that past trusting behavior outside the experiment is significantly correlated with trusting behavior in the experiment.

Our analysis is based on Figures 1 to 3 and Table 4. In Figure 1 we show the distribution of the first-movers' transfers and Figure 2 shows the average transfer of the first movers as a function of the first-movers' beliefs about the transfer of the second mover. Figure 1 indicates that transfers of zero, five and 10 occurred most frequently. 17 percent of the first-movers transferred nothing, 36 percent transferred 5 € and 11 percent transferred 10 €. 24 percent of the participants transferred between 2 and 4 €. Figure 2 shows that the first-movers' transfers are strongly dependent on their beliefs of what they will receive from the second-movers. The higher the believed transfer of the second mover is, the higher the first-movers' transfer is.

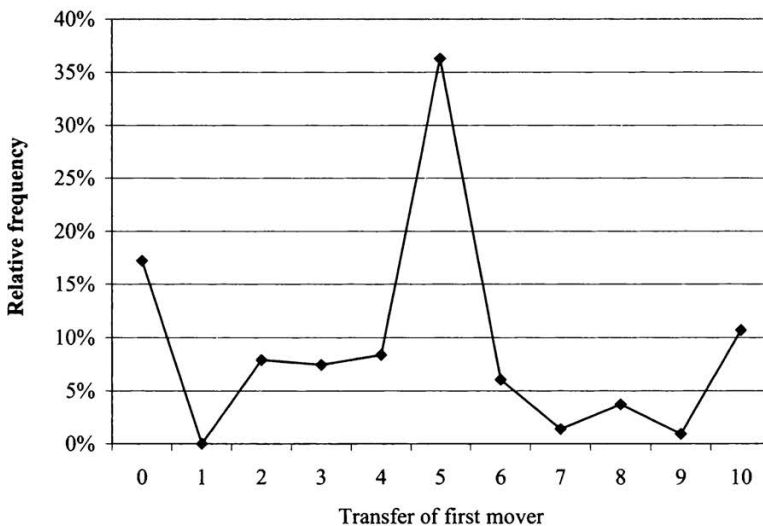


Figure 1: Distribution of first-movers' transfers

In Table 4 we perform ordered probit regressions of first- and second-movers' behavior on a list of demographic variables that we thought are potentially important or that have been suggested in the previous literature on trust and trustworthiness.¹⁵ The ordered probit specification perfectly meets the truncated range of our dependent variable. Regressions (1) and (2) regress the first-movers' transfers on these variables. Model (1) omits the first-movers' beliefs about second-movers' transfers. The results for model (1) show that the older people are, the less trusting they become. The base-

¹⁵ For theoretical considerations behind the variables see Alesino and La Ferrara (2002), Carter and Castillo (2002), and Guiso et al. (2002).

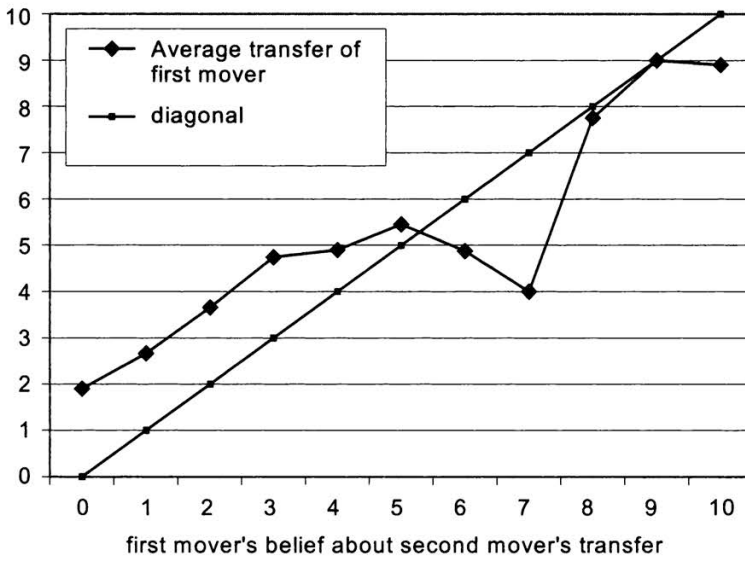


Figure 2: First-movers' transfers as a function of their belief about the second-mover's transfer

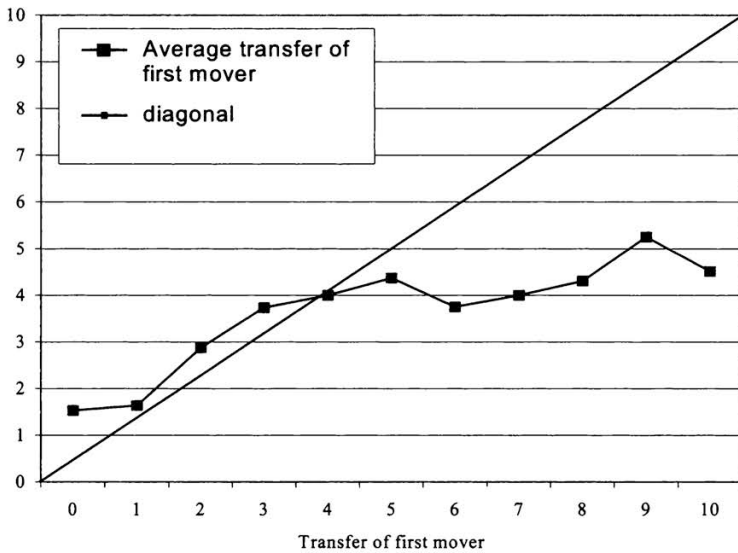


Figure 3: Average transfer of the second movers as a function of the first-mover's transfer

Table 4
**Socio-economic and demographic determinants
of trusting and trustworthy behavior**

	1 st movers transfer (1)	1 st movers transfer (2)	2 nd movers transfer (3)
Female	0.161 (0.173)	0.220 (0.195)	0.170 (0.163)
age below 35	0.106 (0.266)	0.000 (0.285)	-0.437 (0.234)*
age 55 – 65	-0.237 (0.260)	-0.280 (0.294)	0.491 (0.323)
age above 65	-0.505 (0.304)*	-0.849 (0.303)**	0.671 (0.366)*
single child	-0.100 (0.191)*	-0.312 (0.189)*	-0.250 (0.233)
Foreign citizen	0.971 (0.498)*	0.392 (0.463)	-0.691 (0.355)*
Dummy for East Germany	0.228 (0.269)	0.282 (0.256)	-0.087 (0.213)
low education	-0.160 (0.197)	-0.292 (0.195)	-0.131 (0.182)
high education	0.212 (0.227)	0.187 (0.254)	-0.194 (0.339)
high qualification & free professions	-0.503 (0.243)**	-0.101 (0.269)	0.186 (0.380)
size of household	-0.188 (0.101)*	-0.094 (0.118)	0.126 (0.096)
equivalence income	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
works for a living (includes self-employed)	0.032 (0.207)	-0.184 (0.219)	0.357 (0.248)
Unemployed	-0.042 (0.408)	-0.000 (0.443)	-0.653 (0.306)**
Unmarried	-0.181 (0.291)	-0.194 (0.304)	-0.130 (0.216)
rural area (< 5,000 inhabitants)	0.230 (0.232)	0.230 (0.244)	0.087 (0.301)
big city (> 100,000 inhabitants)	0.002 (0.194)	0.087 (0.194)	-0.303 (0.205)
Catholic	0.505 (0.248)**	0.635 (0.258)**	-0.209 (0.270)
Protestant	0.081 (0.214)	0.376 (0.251)	-0.283 (0.203)
goes to church at least once a month	0.037 (0.208)	-0.334 (0.227)	0.144 (0.231)
separated or divorced during last year	-0.350 (0.650)	0.048 (0.804)	-0.844 (0.360)**

Table 4 (continue):

	1 st movers transfer (1)	1 st movers transfer (2)	2 nd movers transfer (3)
worries about own economic situation	0.131 (0.226)	0.166 (0.215)	-0.097 (0.195)
worries much about crime	-0.043 (0.162)	0.013 (0.176)	0.300 (0.180)*
good health state	0.162 (0.167)	-0.173 (0.194)	0.471 (0.187)**
not in favor of any party	-0.067 (0.274)	0.248 (0.252)	-0.443 (0.212)**
in favor of Social Democratic Party	-0.019 (0.299)	0.453 (0.257)*	-0.246 (0.248)
in favor of Christian Democratic Party	0.139 (0.315)	0.660 (0.278)**	-0.767 (0.268)***
belief about transfer of 2 nd mover		0.358 (0.049)***	
received from 1 st mover			0.103 (0.027)***
Observations	213	197	210

Note: Ordered Probit Regressions. Robust standard errors in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%

line age category is the people between age 35 and 55. First-movers who are older than 65 transfer significantly less money, and those who are younger than 35 transfer (insignificantly) more money. Model (1) also shows that foreign citizens and Catholics are significantly more trusting whereas highly skilled participants and individuals living in larger households transfer significantly less money. Whether people live in East or West Germany, whether they have low or high education, live in a big city or in rural areas, or whether they favor the Social Democrats or the Christian Democrats has no significant impact on behavior according to model (1).

In Model (2) we include first-movers' beliefs as a control variable in the regression. In line with Figure 2 the coefficient for beliefs is significantly positive. Not surprisingly the impact of household size, being a highly skilled worker and being a foreign citizen are now insignificant suggesting that the impact of these variables in model (1) operates predominantly via belief formation. However, being older than 65 and being a Catholic remains significant. In addition three other variables become significant. Being an only child decreases first-movers' transfers and being in favor of the Social Democratic Party or the Christian Democratic Party increases transfers.¹⁶

¹⁶ Glaeser et al. (2000) report that only children transfer less when they are in the role of the second-mover but not when they are in the role of the first-mover.

Turning to second-mover behavior, Figure 3 shows that the average transfer of the second-mover is increasing in what they received from first-movers. This is also confirmed by regression model (3), which indicates that the first-movers' transfer is highly significant. In addition model (3) indicates that second-movers below the age of 35 give significantly less whereas those above 65 give significantly more than the comparison group aged 35–54. Foreign citizens, unemployed workers, people who recently separated from their (marriage) partner, people who indicate that they favor none of the existing parties, and people who indicate that they favor the Christian Democratic Party give less. These party effects, however seem to be driven by older respondents because they vanish if the sample is restricted to respondents below age 60 (not displayed in a Table). Finally, second-movers with good health tend to give significantly more.

Summing up it is interesting that gender, people's income situation (corrected for household size), worries about the own economic situation and whether people have comparatively low or high education does not affect either first- or second-mover behavior.

We checked the above results in several ways. We conducted, e.g., regressions restricted to those below the age of 60 to check whether the unemployed still give less when the sample is constrained in this way. The rationale for this is that the unemployment variable in model (3) might also pick up behavioural differences between the unemployed and the retired. The negatively significant impact of being unemployed on the second-movers transfer is, however, robust. The main effect of excluding respondents above the age of 60 was that all the party effects in model (2) and (3) vanished.

Due to the small number of cases influential cases that dominate the results of the regressions are very likely. But for the ordered probit regression there is no straightforward tool at hand to aid in the search for influential cases. So we calculated the leverage (diagonal elements of the projection hat matrix) using an OLS regression for detecting those cases. Applying OLS to our data is not ideal but not wrong because our dependent variable has 11 values which is – in practical terms – a sufficient range for OLS. We found between 4 and 6 influential cases in our four models. After taking them out of the ordered probit regressions the coefficients did not change a lot. But there was one exception: due to the small number of cases we lost the observations for separated respondents and thus the respective significant effect.

In our view the most interesting results from this preliminary analysis concerns the age effects, the impact of high skills, household size and foreigner status on beliefs (and hence first-mover transfers) and the impact of health status and unemployment on second-mover behaviour. The latter ef-

fects suggest that satisfaction with life and jobs may also have important effects on trust and trustworthiness – a question we intend to explore in future work. While researchers typically concentrate on significant effects we believe that it is also interesting to focus on those variables, which did not affect behavior in our small sample. For instance, income (adjusted for household size) and worries about one's own economic situation have no significant impact. Likewise, living in big cities or in rural areas leaves trust and trustworthiness unaffected. It is, of course, always possible that with more cases these variables eventually become significant. However, in our sample these variables are far from being significant which does not raise confidence that they eventually will turn out to be significant as sample size increases.

6. Conclusions

In this paper we proposed to combine representative surveys with representative experiments. This combination enables researchers to overcome important limits of both methods and cross-validate survey data and behavioural data from experiments. However, experiments involving the sequential interaction between participants who are separated in subsamples of first and second movers pose a methodological challenge because a survey consists of independent interviews of non-interacting respondents. We proposed to solve this problem by starting from an empirically plausible ex-ante distribution of first-mover choices. Second-movers in the experiment are confronted with randomly chosen first-mover decisions from the ex-ante distribution. When all first- and second mover decisions have been collected the ex-post distribution of first-mover choices is used to assign to each second-mover who was confronted with a first-mover decision x a first-mover who actually has chosen x . With the exception of a few marginal cases such ex-post matching is possible if the ex-ante distribution is based on reliable empirical information about the likely ex-post distribution of first-mover decisions (and overshooting cases can be matched in a fair manner as well). The big advantage of this method is that it integrates experiments that require social interaction between respondents into surveys resting on independent interviews in a smooth and inexpensive way.

We illustrated the power of our approach with the help of a representative social dilemma experiment designed to measure respondents trust and trustworthiness. We identified those survey questions on trust and trustworthiness that are the best predictors of behaviorally exhibited trust. This information is quite useful for future survey research. It turns out that questions about past trusting behavior and direct questions about people's trust are the best predictors of trusting behavior. Our approach also allowed us to

identify important socio-economic determinants of trust and trustworthiness. Retired people, high skilled workers and respondents from larger households trust less, whereas Catholics and foreign citizens exhibit more trusting behavior. Retired people and healthy people show more trustworthy behavior whereas those below the age of 35, foreign citizens, and the unemployed exhibit less trustworthy behavior. Although our analysis is still at a preliminary stage we regard this as promising results that encourage further investigation by means of large data sets.

References

- Alesino, A. / La Ferrara, E.* (2002). Who trusts others? *Journal of Public Economics* 85: 207 – 234.
- Altonji, J. G. / Devereux, P. J.* (1999). The Extent and Consequences of Downward Nominal Wage Rigidity, NBER Working Paper # 7236.
- Berg, J. / Dickhaut, J. / McCabe, K.* (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10: 122 – 142.
- Camerer, C. / Hogarth, R. M.* (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework, *Journal of Risk and Uncertainty* 19, 7 – 42.
- Carter, M. R. / Castillo, M.* (2002). The economic Impacts of Altruism, Trust and Reciprocity: An Experimental Approach to Social Capital. Working Paper No. 448, University of Wisconsin.
- Cooper, D. / Kagel, J. H. / Lo, W. / Liang Gu, Q.* (1999). Gaming against Managers in Incentive Systems: Experiments with Chinese Managers and Chinese Students, *American Economic Review* 89: 781 – 804.
- Fehr, E. / Goette, L.* (2000). Robustness and Real Consequences of Nominal Wage Rigidity, University of Zürich, Institute for Empirical Research in Economics, Working Paper No. 44.
- Fehr, E. / Kirchsteiger, G. / Riedl, A.* (1993). Does Fairness Prevent Market Clearing? An Experimental Investigation. *Quarterly Journal of Economics* 58: 437 – 460.
- Fowler, F. J.* (1988). *Survey Research Methods*. Newbury Park, London and New Delhi.
- Glaeser, E. L. / Laibson, D. I. / Scheinkman, J. A. / Soutter, C.* (2000), Measuring Trust, *Quarterly Journal of Economics* 115: 811 – 846.
- Guiso, L. / Sapienza, P. / Zingales, L.* (2002). People's Opium? Religion and Economic Attitudes. CEPR Discussion Paper No. 3588. London.
- Knack, St. / Keefer, Ph.* (1997), Does Social Capital have an Economic Payoff? A Cross-Country Investigation, *Quarterly Journal of Economics* 112, 1251 – 1288.
- Schupp, J. / Wagner, G. G.* (2002). Maintenance of and innovation in long-term panel studies: the case of the German Socio-Economic Panel (SOEP). *Allgemeines Statistisches Archiv* 86, 163 – 175.

Appendix 1

Sampling Method to generate the Random Sample for our Survey-Experiment

Our random sample is based on the standard sampling method generated by the association of private market and social research agencies in Germany (ADM-Stichprobensystem). For each of 100 randomly chosen PSU (primary sampling units, i.e. *Wahlbezirke*) one experienced and trained interviewer was given a randomly chosen starting address. Starting at that specific local address¹ the interviewer had to contact every third household and had to motivate one adult person age 16 and older to participate. Therefore, only one person per household participated in our survey-experiment. 53 percent of the households contacted were willing to participate in the survey.² This participation rate is typical for randomly sampled surveys in Germany.

Each interviewer had to recruit 5 interviews per sample point. All recruited respondents in a given PSU were assigned the same role in the sequential dilemma game. All respondents in a PSU were either all in the role of player A (first mover) or all were in the role of player B (second mover). This method allowed a balanced split between the two subsamples defined by the role of the players in the game. 50 interviewers surveyed only participants in the role of player A and 50 interviewers surveyed participants in the role of player B. In addition, since the interaction between respondents took place only across PSU's the anonymity of the respondents was guaranteed. Even the interviewer did not know with whom a particular participant was matched.

¹ Based on the specific standard instructions for a random walk.

² The interviewer could not reach about half of all nonrespondents within the very short field period.