DISCUSSION PAPER SERIES

IZA DP No. 6916

Trust, Values and False Consensus

Jeffrey Butler
Paola Giuliano
Luigi Guiso

October 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Trust, Values and False Consensus

Jeffrey Butler
EIEF

Paola Giuliano
UCLA, NBER, CEPR and IZA

Luigi Guiso
EIEF and CEPR

Discussion Paper No. 6916
October 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 6916
October 2012

ABSTRACT

Trust, Values and False Consensus*

Trust beliefs are heterogeneous across individuals and, at the same time, persistent across generations. We investigate one mechanism yielding these dual patterns: false consensus. In the context of a trust game experiment, we show that individuals extrapolate from their own type when forming trust beliefs about the same pool of potential partners – i.e., more (less) trustworthy individuals form more optimistic (pessimistic) trust beliefs - and that this tendency continues to color trust beliefs after several rounds of game-play. Moreover, we show that one's own type/trustworthiness can be traced back to the values parents transmit to their children during their upbringing. In a second closely-related experiment, we show the economic impact of mis-calibrated trust beliefs stemming from false consensus. Mis-calibrated beliefs lower participants' experimental trust game earnings by about 20 percent on average.

JEL Classification: A1, A12, D1, Z1

Keywords: trust, trustworthiness, culture, false consensus

Corresponding author:

Paola Giuliano
Anderson School of Management
UCLA
110 Westwood Plaza
Entrepreneurs Hall C517
Los Angeles, CA 90095
USA
E-mail: paola.giuliano@anderson.ucla.edu

* We are grateful to seminar participants at the Bank of Spain, the California Center for Population Research at UCLA, the Einaudi Institute for Economics and Finance, the Kaler Meeting at UCLA, the FEEM conference on the Economics of Culture, Institutions and Crime, the EALE/SOLE joint conference in London, the 9th IZA/SOLE Transatlantic Meeting of Labor Economists, the seventh International Meeting on Behavioral and Experimental Economics, the Higher School of Economics in Moscow, the London School of Economics, the University of California Davis, the NBER Political Economy Meeting, University of Mannheim, Universidad Pompeu Fabra, University of San Diego, University of Siena, Stanford University and Toulouse University for helpful comments. Luigi Guiso thanks EIEF for financial support and LUISS University for making the LUISS lab available. Paola Giuliano thanks the UCLA-CIBER grant for financial support.

1 Introduction

A large strand of literature has shown the persistence of trust beliefs across generations, using evidence from different datasets and a variety of countries.[1] Trust beliefs are at the same time also quite heterogeneous across individuals[2]. In this paper we provide evidence suggesting that false consensus, the tendency of individuals to extrapolate the behavior of others from their own type (Ross, Green and House,1977), may be able to explain these dual patterns.

Persistent heterogeneity in trust beliefs, even in the same community, has been explained in literature in various ways. According to one view, individuals' beliefs are initially acquired through cultural transmission and then slowly updated through experience from one generation to the next. This line of argument has been pursued by Guiso, Sapienza and Zingales (2008b) who build an overlapping-generations model in which children absorb their trust priors from their parents and then, after experiencing the real world, transmit their (updated) beliefs to their own children. Dohmen et. al (2012) provide evidence consistent with this view. Heterogeneity is the result of family specific shocks. Within a generation, correlation between current beliefs and received priors is diluted as people age and learn. Yet this dilution needs not to be complete and a high degree of persistence may still obtain.

On the other hand, a slightly different explanation is that parents instill values, such as trustworthiness, rather than beliefs. Cultural transmission of values of cooperation and trustworthiness is the focus of Bisin and Verdier (2000), Bisin, Topa, and Verdier (2004) and Tabellini (2008). They show how norms of behavior are optimally passed down from parents to children

[1] See Algan and Cahuc (2010), Butler, Giuliano and Guiso (2012), Dohmen et al. (2012), Guiso, Sapienza and Zingales (2008a).

[2] Butler et al. (2012) and Dohmen et al. (2012)

and persist from generation to generation. Heterogeneity in parents' preferences and experiences may then result in heterogeneity in instilled trustworthiness. Even if parents do not teach beliefs directly, individuals may extrapolate from their own type when forming beliefs about others' trustworthiness. As Thomas Schelling once wrote "you can sit in your armchair and try to predict how people behave by asking yourself how you would behave if you had your wits about you. You get free of charge a lot of vicarious empirical behavior" (1966, p. 150).

In this paper we show that false consensus is a mechanisms that could help to explain how heterogeneity in values could translate into heterogeneity in beliefs. We view false consensus as a source of initial prior. In the absence of a history of information about the reliability of a pool of people, those interacting with an unknown pool form a prior by asking themselves how they would behave in similar circumstances. Since they would behave differently, they start with different priors. If values (or priors) persist over time and false consensus does not vanish with learning, then wrong beliefs will also persist. In our context false consensus implies that highly trustworthy individuals will tend to think that others are like them and form overly optimistic trust beliefs, while highly untrustworthy people will extrapolate from their own type and form excessively pessimistic beliefs. Both highly trustworthy and highly untrustworthy individuals will tend to systematically form more extreme trust beliefs than are warranted by their experiences. A long history of research on false consensus has indeed shown it to be a persistent phenomenon (Krueger and Clement (1994)) that need not drowned out by monetary incentives for accurate predictions (e.g. Massey and Thaler (2006)).

To show the relevance of false consensus we conduct two experiments. The first experiment implements a repeated version of the standard trust

game in the laboratory (Berg, Dickhaut and McCabe, 1995). The experiment allows us to obtain a measure of participants' own (initial) trustworthiness and also to elicit participants' beliefs after each round of game play. We first document a strikingly high correlation between participants' trust beliefs and their own trustworthiness, suggesting that beliefs are formed by extrapolating from one's own type. Moreover, we show that this correlation remains strong and significant even after several rounds of game play. In addition, we also investigate where individual priors are coming from, showing that initial trustworthiness can be traced back to the values instilled by our participants' parents during their upbringing.

In a second experiment, we investigate the economic consequences of false consensus. We show that it is indeed the case that the most (least) trustworthy participants tend to form overly-optimistic (overly-pessimistic) trust beliefs and, consequently, trust more (less) than they should. Participants with miscalibrated beliefs earn in the process 18% less than participants with properly-calibrated beliefs.

The remainder of the paper proceeds as follows. In Section 2, we describe the experiment aimed at showing the relevance of false consensus and its persistence. In Section 3, we present the design of Experiment 2 and show the results about the economic costs of false consensus. In section 4, we summarize our findings and present concluding remarks.

2 False consensus, values and persistence

2.1 Experiment 1: design and procedures

Participants were recruited from a pre-existing list of students who had previously expressed willingness to take part in experiments, in general, at LUISS Guido Carli University in Rome, Italy. All laboratory sessions were conducted at CESARE, the lab facility at LUISS. The experiment

was programmed and implemented using the software z-Tree (Fischbacher, 2007). In total, 124 students participated in Experiment 1.

After showing up to the lab at pre-scheduled session times, participants were seated at individual desks in the lab each equipped with its own computer. Participants were separated from one another by opaque dividers. Once all participants were seated, instructions were read aloud and participants' questions, if any, were answered by the experimenters.

After instructions were read and questions were answered, subjects proceeded to the game-playing phase. This phase consisted of up to twelve rounds of the trust game, as described below. Participants were not informed how many rounds of game-play there would be, but rather only instructed that there would be "several" rounds. This was meant to minimize end-game effects possible when the number of rounds is known. Because sessions were scheduled to last (up to) two hours, and because most participants had never participated in any experiment before (CESARE is a new facility) the number of rounds per session varied widely. Sessions consisted of anywhere from 3 to 12 rounds, with the majority consisting of 12 rounds.

Before each round, each participant was randomly and anonymously (re-) matched with a co-player, and within each resulting pairing roles were randomly (re-)assigned. These design features allow for learning about the population's traits and preferences but not about any specific person's traits/preferences. They also serve to ameliorate many repeated-game effects that are possible when partners are uniquely identifiable or persist over rounds, such as reputation building or directly punishing/rewarding specific partners for past behavior, as such effects are not the focus of this experiment.

The trust game is a two-player sequential-moves game of perfect information. The first-mover ("sender") is endowed with 10.50 euros. The second-

mover—the "receiver"—is given no endowment. The sender chooses to send some, all or none of his or her endowment to the receiver. Any amount sent is tripled by the experimenter before being allocated to the receiver. The receiver then chooses to return some, all or none of this tripled amount back to the sender, ending the game. Sending a positive amount entailed a small fee—0.50 euros.

Feasible actions for the sender in our implementation were to send any whole-euro amount: $0, 1, 2, \ldots, 10$. Receivers' decisions were collected using the strategy method. Before receivers discovered how much their sender sent, they specified how much they would return for any amount of money they could receive. One critique of the strategy method is that it is "cold" and does not elicit the same reaction as if participants are faced with an actual decision. To (partially) address this critique, and make receivers' decisions feel as real as possible, receivers were faced with a series of ten separate screens. Each screen asked only one question: "if you receive m euros, how much will you return?" For each separate screen, m was replaced with exactly one value, $m \in \{3, \ldots, 30\} = \{3 \times 1, \ldots, 3 \times 10\}$. The order of possible amounts, m, was randomized in order to avoid inducing any artificial consistency in receivers' strategies. This random order was the same for all receivers within each round, and was re-randomized between rounds. Obviously, no information about receivers' decisions was shared with senders in any way before the end of each round.

At the end of each round, each sender and receiver pair was informed of the outcome of their interaction—i.e., how much the sender sent, and, if this was a positive amount, how much the receiver returned as determined by the relevant element of the receiver's strategy vector. No other element of the receiver's strategy vector was revealed, nor was any information about the outcome in any other participant pair.

To collect beliefs, within each round every participant—regardless of the role they had been assigned—was asked to estimate the amounts receivers would return, on average, for each possible amount receivers could receive. Specifically, participants answered ten questions: "How much will receivers return, on average, if they receive m euros?", $m \in \{3, \ldots, 30\}$. Participants who were currently receivers were told to exclude their own actions from this estimate and that they would be remunerated on this basis. That is to say, they were asked to estimate how much *other* receivers would return. This serves to rule out any mechanical—real or imagined—connection between participants' own actions and their estimates.

Incentives to report beliefs truthfully were given by paying subjects according to a quadratic scoring rule[3]. Beliefs were elicited either before or after participants submitted their actions, with this order being randomly re-determined for each participant before each round.

When all rounds were completed, one round was selected at random and participants were paid in accordance with their actions and the accuracy of their estimates in that round. This procedure is meant to eliminate wealth effects from accumulated earnings over rounds and is standard in the literat-

[3]It is well-known that this rule gives (risk-neutral) individuals incentives compatible with reporting truthfully the mean of their subjective distribution of beliefs. Specifically, for each of the ten belief questions participants earned an amount of money given by the function below, where $\widehat{r_m}$ is receivers estimated return amount, r_m is receivers actual (average) return amount, and as above $m \in \{3, \ldots, 30\}$:

$$Earnings = 1 - (\frac{\widehat{r_m} - r_m}{m})^2$$

For example, if a subject's estimate of receivers' average return amount, conditional on receiving 9 euros, was 6 euros–i.e., $\widehat{r_9} = 6$—and receivers' strategy vectors entailed returning (on average) 2 euros conditional on receiving 9, then that participant's estimate would earn the participant (in euros)

$$1 - (\frac{6-2}{9})^2 = 1 - \frac{16}{81} \approx 0.80 \tag{1}$$

A perfect estimate paid 1 euro, so that subjects could earn up to 10 euros each round from their estimates.

ure. All of these design elements were (commonly) known by all participants.

2.2 Our uni-dimensional measures of trust beliefs and trustworthiness

To construct a unidimensional measure of trust beliefs for each participant, we converted each of the 10 elements of his or her belief vector into percentage terms (0 to 1) and then took the average of these ten percentages. For example, suppose a participant's belief vector is $(1, 2, \ldots, 10)$—i.e., they believe that receivers will on average return 1 if they receive $3 \times 1 = 3$, 2 if they receive $3 \times 2 = 6$, etc. We divide the first element by 3, the second by 6 and so on, to get the modified belief vector $(\frac{1}{3}, \frac{2}{6}, \ldots, \frac{10}{30})$ and then average over the elements of this vector to get $\frac{1}{3}$, or 0.33, as the participant's uni-dimensional trust belief. To get a unidimensional measure of trustworthiness, for each receiver we apply the same procedure to their willingness-to-return vector. Consequently, we obtain a uni-dimensional trust belief measure for all participants, and a unidimensional trustworthiness measure for half of the participants for each round of game-play - those assigned the role of receiver.

As a measure of "initial" trustworthiness largely untainted by learning, we assign to each individual their unidimensional trustworthiness measure from the first time they played receiver, provided this occurred in one of the first two rounds.[4] Since roles are randomly re-assigned each round, this measure is defined for a large majority of participants, but not all of them (92 of 124).

[4] The choice of the first two rounds balances two concerns: i) contamination by learning which suggests only including those who were receivers in the first round—and leaving the measure undefined for half of the participants; ii) concerns about sample size which suggest extending the definition to include as many rounds as possible. In the end, we believe our definition is reasonable.

2.3 Parentally-instilled values

Finally, all participants filled out a brief survey. The survey was sent (e-mailed) several days removed from laboratory sessions—a week before or after the participant's session—to mitigate concerns that participants' survey responses could systematically affect their decisions in the lab. One part of the survey asked respondents to report, on a scale from 0 to 10, how much emphasis their parents placed on a number of principles and behavioral rules during their upbringing (frugality, prudence, loyalty, etc.).[5] We use answers from a subset of these questions to construct a measure of the strength of received cultural values and norms of trustworthiness for each participant.

2.4 Results

Figure 1 shows the distribution of (uni-dimensional) trust beliefs in the first round of the trust game, when no learning about the trustworthiness of the pool of participants had yet been possible (panel A) and of our behavioral measure of own initial trustworthiness (panel B). Since trust beliefs and trustworthiness are measured by the average share that participants expect receivers will send back, and by the average share that receivers are willing to send back, respectively, these variables take values between 0 and 1. As these measures are continuous variables we report kernel density estimates. The figure documents considerable heterogeneity in trust priors. Since beliefs in the experiment refer to a common pool of people, heterogeneity in trust beliefs cannot be automatically ascribed to variation in the pools of people whose trustworthiness is being estimated.[6] Furthermore, since beliefs are

[5] A wide array of questions was asked, some completely irrelevant to trust and trustworthiness, in order to mitigate experimenter/demand effect in the survey answers and in the experiment.

[6] It is true that Figure 1, panel A, reports beliefs for all sessions pooled, so some people might still question the source of heterogeneity. However, plotting the trust belief densities for each session separately (not reported, but available upon request) also yields quite a lot of heterogeneity.

measured independently of behavior, the heterogeneity in Figure 1, panel A, cannot reflect differences in risk attitudes.[7] In the sample the average level of trust beliefs is 0.27 and the sample standard deviation is 0.16.[8]

The figure also documents substantial heterogeneity in behavioral trustworthiness, whose sample mean and standard deviation are 0.32 and 0.16, respectively. In the next section we test whether heterogeneity in trustworthiness is reflected in heterogeneous beliefs.

Table 2, panel A, shows regressions of trust beliefs in various rounds on own initial trustworthiness. To isolate, as best as possible, trustworthiness as an individual trait, we use initial trustworthiness as a regressor. To reduce sampling variation due to small sample size we aggregate observations over blocks of three rounds. As the first column shows, in early rounds initial trustworthiness is strongly positively correlated with trust beliefs, lending support to the idea that individuals form beliefs about others' trustworthiness by extrapolating from their own types. Quite remarkably, own trustworthiness explains about 60% of the initial heterogeneity in beliefs. As the second column shows, this tendency does not vanish when the game is repeated and people are thus given the opportunity to learn about the pool of participants. The correlation weakens, and the effect is somewhat smaller, in later rounds but both remain sizable and significant. Thus, initial trustworthiness still affects trust beliefs even after the game has been played several times, always drawing from an invariant pool of individuals, which we take as evidence that false consensus persists. However, the decline in the strength of the link also suggests that given enough opportunities to

[7] Unless the elicitation procedure is biased by risk preferences as well. We cannot rule this out completely, as how to do so is a still-unsettled debate within experimental economics. We use a very standard quadratic scoring rule. There is experimental evidence suggesting that this mechanism elicits beliefs reasonably accurately regardless of risk preferences (see, e.g., Huck and Weiszäcker, 2002).

[8] Since every dollar sent is tripled, 0.33 would imply senders believe that receivers will return as much as is sent.

learn about a stable pool of people, the tendency to attribute to others one's own trustworthiness may vanish.[9]

This evidence is consistent with the idea that priors are driven, through false consensus, by norms of behavior that shape individual's own trustworthiness. To make this link even more clear and show the ultimate relationship between cultural values and beliefs we use information on the moral values emphasized by participants' parents. For our purposes, we use parents' emphasis on two values: the first is how much emphasis an individual's parents placed on teaching to always behave as good citizens; the second is the emphasis parents placed on loyalty to groups or organizations. We average the responses to these two questions and divide the result by 10 to put it on a scale—0 to 1—comparable with beliefs. We use this measure as a proxy for individuals' intrinsic trustworthiness, an individual-specific trait.

Table 2, Panel B shows that this measure of parents' effort spent on teaching good values is correlated with individuals' initial trustworthiness, which is consistent with behavioral types reflecting heterogeneous cultural values.[10] Of course, it is imperfectly correlated, partly because the measure of values that we have is only a proxy for the true trait, and partly because own traits are also shaped by interactions in the social sphere (i.e. through socialization). Panel C shows direct regressions of trust beliefs on our sur-

[9] An interesting question is whether the false consensus effect reappears any time an individual faces a new pool of people or the pool she is interacting with changes.

[10] One might worry that this correlation simply reflects priming participants to think about morality by the mere fact of answering the survey. If so, one would expect the correlation to be particularly strong for participants who took the survey *before* their experimental session. We check for this by splitting the sample into those who took the survey before their session and those who took the survey after their session. The correlation between good values and initial trustworthiness is positive in both subsamples, but is larger in the subsample of those who took the survey after the experiment. As a second check, we inserted a dummy into the simple univariate regression of trustworthiness on values (not reported, but available upon request) that takes the value of one if a participant took the survey after the experiment. The coefficient on this dummy is non-significant, there is very little change in the coefficient on good values (it falls slightly to 0.159) and there is no change in the significance level of this coefficient.

vey measure of cultural values: at all repetitions the cultural measure of trustworthiness predicts trust beliefs.

In sum, the evidence from Experiment 1 shows three things. First, when no information is available about a group, individuals form beliefs about the trustworthiness of others extrapolating from their own types, which are quite heterogeneous. Second, this tendency is highly persistent, though attenuated through learning. Third, heterogeneity in own trustworthiness can be traced back to heterogeneous cultural norms instilled by parents implying that measures of the latter can provide valuable instruments for trust beliefs, an implication which could prove useful in empirical investigations of trust beliefs.

3 The economic costs of false consensus

3.1 Experiment 2: design and procedures

Participants were recruited from the same pre-existing list of potential experimental student participants at LUISS in Rome, Italy. All sessions were conducted on-line. This experiment was conducted on four separate days, each day constituting a session. In total, 122 students participated in the on-line experiment. We excluded from the list of invitees anybody who had taken part in the laboratory experiment, so that no individual took part in both the in-lab and the on-line experiment.

The on-line experiment implemented one round of the trust game in the same manner as above with three exceptions. The first exception is that the function used to transform money sent into money received was no longer linear, but rather quadratic. This function was presented to participants in table form (below). Using a quadratic "trust production function" will aid us in the investigation of the intensive margin of trust as it provides an internal optimal send amount for a wide array of trust beliefs and preferences where

a linear function would yield corner solutions. Secondly, a full strategy method was used: participants submitted their decisions in both possible roles before learning which role they would be assigned. Finally, participants did not know their beliefs would be elicited until after they submitted their decisions. This weakens concerns that belief elicitation itself could affect decisions.

If the sender sends (euros):									
1	2	3	4	5	6	7	8	9	10
Then the receiver will receive (euros):									
8.05	11.30	13.85	16.05	17.90	19.60	21.20	22.65	24.05	25.30

In terms of earnings, two features are notable. First, belief accuracy was remunerated using a slightly different procedure: a randomized quadratic scoring rule. Schlag and van der Weele (2009), among others, have shown that this procedure is theoretically robust to individual risk preferences. Specifically, each estimate is converted into a number, $z \in [0, 1]$, precisely as above. At the same time, the computer chooses at random a number, $y \in [0, 1]$. If $y \leq z$, the participant earns 5 euros, otherwise the estimate pays nothing. At the end of the session, one estimate is randomly chosen to count towards a participants potential earnings. This latter feature should allay concerns about hedging across belief estimates that would be possible if, as in Experiment 1, all ten estimates were remunerated with certainty. The second feature of note is that only 10 percent of participant pairs were (randomly) chosen to be paid according to their decisions and estimates. Since the on-line experiment required much less of participants' time, this kept hourly earnings comparable to earnings in the laboratory experiment.

For our analysis we make use of a uni-dimensional measure of trust beliefs and trustworthiness obtained using the same procedure as in Experiment 1 (described above). Since we here use a full strategy method, however, we

have both measures for all participants.

Next, for each participant, i, we construct a measure of performance by randomly choosing another participant, j, from the same experimental session and computing i's earnings using i's sender strategy and j's receiver strategy.[11]

Finally, we use willingness-to-return amounts—excluding each participant's own actions—and beliefs about these return amounts within each session to construct a unidimensional measure of belief errors for each participant. Specifically, for each participant we first compute a separate belief error in percentage terms for each amount a receiver could have received. This yields ten separate belief error measures for each participant, each ranging from -1 to 1, where negative values indicate under-estimating. We use the average of these ten measures for each participant as our uni-dimensional belief errors measure, which again ranges from -1 to 1.

3.2 Results

Figure 2 presents a scatter plot of the relationship between our belief errors measure and performance in Experiment 2. We find evidence for false consensus again: belief errors are positively correlated with own trustworthiness ($\rho = 0.39; p < 0.01$;a scatter plot of belief errors and own trustworthiness is presented in Figure 2.). We also find that earnings are hump-shaped in belief errors. Both those who hold overly pessimistic trust beliefs (negative belief errors) and those who hold overly-optimistic trust beliefs (positive belief errors) earn less than those whose belief errors are approximately zero. This humped shape is confirmed by the regression presented in Table 3: the

[11] That is, performance for participant i is measured as the earnings they would have made if they had been assigned the role of sender: $Y_i = 10.5 - S_i + \gamma_j 8 S_i^{0.5} - 0.5 I(S_i)$, where γ_j denotes the proportion of the amount received, $8S_i^{0.5}$, what the receiver j paired with i returns and $I(S_i)$ is an indicator function equal to 1 if i sends a positive amount.

coefficient on the squared belief errors is both negative and significant.[12] Furthermore, the coefficient on the linear term, regardless of significance, implies that performance attains its maximum for belief errors close to zero. The estimated relationships suggest that senders earn between 11.00 and 11.45 euros on average when belief errors are zero, constituting a 5 to 9 percent increase over the safe return (10.50 euros) from sending nothing.[13]

To get another measure of the magnitude of income differences implied by belief errors we divided the data into three categories: "under-estimators," "over-estimators," and "accurate-estimators." Accurate-estimators had belief errors within a small interval around zero, $[-0.1, 0.1]$; under-estimators had belief errors below this interval; over-estimators had belief errors above this interval. Table 4 shows that accurate-estimators earned about 18 percent more on average than under-estimators, who, in turn, earned about the same as over-estimators.[14]

Summing up, Experiment 2 allows us to investigate the economic consequences of false consensus. Consistent with false consensus, we find that own trustworthiness colors trust beliefs and that this has a significant pecu-

[12] This continues to be true when we add dummies for each session to control for session fixed effects and when standard errors are clustered by session, where each separate day the experiment was conducted constitutes a session.

[13] One potential concern common to most experimental research relates to stake size. It could be that participants rely on heuristics such as extrapolating from their own types only when stake sizes are small. Although we cannot directly address that concern here since we did not vary the payoffs for correct beliefs in this experiment, we have a related paper which uses the same "quadratic trust game" in which we vary payoffs for correct beliefs across sessions (Butler, Giuliano and Guiso, 2012). There, in some treatments exactly correct beliefs pay 5 euros—as they do here—while, in other treatments, exactly correct beliefs earn the paricipant four times as much—20 euros. We find that the correlation between own trustworthiness and beliefs *increases* when the payment for correct beliefs increases.

[14] As rough robustness checks (not reported, but available on request) we also ran the regressions in Table 4 using a wider interval—$[-0.15, 0.15]$—or a narrower interval—$[-0.05, 0.05]$—to define accurate-estimators, as well as using a definition of over- and under-estimators defined by the 33rd and 66th percentiles of the observed belief errors. None of these modifications change the results qualitatively: accurate-estimators consistently earned more, on average, than others.

niary impact as the resulting mis-calibrated trust beliefs reduce earnings in our experiment by roughly 20 percent, on average.

4 Conclusions

Large-scale survey evidence suggests that trust beliefs are both extremely heterogeneous across individuals and persistent over age and across generations. In this paper we present the results of two experiments aimed at investigating one prevalent phenomenon that can explain both of these patterns: false consensus. We show that individuals extrapolate from their own type when forming trust beliefs about a novel population (false consensus) and that one's own type continues to have a substantial impact on trust beliefs even after considerable opportunities for learning about the population. In our second experiment we use a trust game slightly modified to allow behavioral trust to more smoothly vary with trust beliefs than in the canonical game. This permits us to investigate how false consensus may hinder earnings. In this one-shot setting, we again find evidence for a substantial impact of false consensus: mis-calibrated trust beliefs stemming from false consensus lower participants' earnings by 20 percent, on average.

References

[1] Algan, Yann and Pierre Cahuc (2010), "Inherited Trust and Growth," *American Economic Review*, 100(5): 2060-92.

[2] Berg, J., Dickhaut, J. and K. McCabe (1995), "Trust, Reciprocity and Social History," *Games and Economic Behavior*, 10, 122-142.

[3] Bisin, Alberto, Giorgio Topa, and Thierry Verdier (2004). "Cooperation as a Transmitted Cultural Trait," *Rationality and Society,* 16 (4), 477-507.

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

[5] Butler, Jeffrey V., Paola Giuliano and Luigi Guiso (2012), "The Right Amount of Trust," EIEF Working Paper.

[6] Butler, Jeffrey V., Paola Giuliano and Luigi Guiso (2012), "Cheating in the Trust Game," EIEF Working Paper.

[7] Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2012). "The Intergenerational Transmission of Risk and Trust Attitudes," *The Review of Economic Studies*, 79(2), 645-677.

[8] Guiso, Luigi, Paola Sapienza and Luigi Zingales (2008a), "Long Term Persistence," NBER WP 14278.

[9] Guiso, Luigi, Paola Sapienza and Luigi Zingales (2008b), "Social Capital as Good Culture," *Journal of the European Economic Association*, 6(2–3), 295–320.

[10] Krueger, Joachim and Russel W. Clement (1994), "The Truly False Consensus Effect: An Ineradicable and Egocentric Bias in Social Perception", *Journal of Personality and Social Psychology of Addictive Behaviors*, 67 (4):596-610.

[11] Massey, Cade and Richard H. Thaler (2006), "The Loser's Curse: Overconfidence vs. Market Efficiency in the National Football League Draft," University of Chicago, mimeo.

[12] Ross, Lee, Greene, D., and House, P. (1977), "The False Consensus Phenomenon: An Attributional Bias in Self-Perception and Social Perception Processes," *Journal of Experimental Social Psychology*, 13(3), 279-301.

[13] Tabellini, Guido (2008). "The Scope of Cooperation: Values and Incentives," *Quarterly Journal of Economics*, 123 (3), 905–950.

Table 1
Descriptive statistics

A. Experiment 1

Variable	mean	St dev
Good Values	0.637	0.199
Initial own trustworthiness	0.32	0.162
Expected trustworthiness (trust belief)	0.265	0.158
Return Proportion	0.211	0.18
Invest Amount	5.258	3.107
Invest Propensity	0.676	0.469

B. On-line experiment

Variable	mean	St dev
Invest Propensity	0.730	0.446
Invest Amount	3.934	3.315
Estimates of Return Proportion	1.287	0.578
Return Proportion	1.312	0.669
Trust Belief Error	-0.007	0.145
Sender Earnings	10.950	3.077

Table 2
The effect of own trustworthiness on trust beliefs

A. OLS estimates of expected trustworthiness on own initial trustworthiness

	Rounds 1-3 Expected trustworthiness	Rounds 4-6 Expected trustworthiness	Rounds 7-9 Expected trustworthiness	Rounds 10-12 Expected trustworthiness
Own initial trustworthiness	0.744***	0.542***	0.475***	0.452***
	(0.0419)	(0.0652)	(0.0748)	(0.0766)
Constant	0.0848***	0.106***	0.0763***	0.0653**
	(0.0161)	(0.0232)	(0.0264)	(0.0246)
Observations	276	208	171	171
R-squared	0.586	0.312	0.261	0.249

B. OLS estimate of initial trustworthiness on "good values"

	Initial trustworthiness
Good Values	0.169*
	(0.0928)
Constant	0.211***
	(0.0597)
Observations	83
R-squared	0.039

C. OLS estimates of expected trustworthiness on good values

	Rounds 1-3 Expected trustworthiness	Rounds 4-6 Expected trustworthiness	Rounds 7-9 Expected trustworthiness	Rounds 10-12 Expected trustworthiness
Good Values	0.122**	0.125*	0.122*	0.0515
	(0.0588)	(0.0662)	(0.0725)	(0.0824)
Constant	0.246***	0.197***	0.143***	0.171***
	(0.0376)	(0.0434)	(0.0448)	(0.0531)
Observations	339	262	216	216
R-squared	0.025	0.027	0.027	0.004

Notes: [1] Robust standard errors, clustered by participant, are reported in parentheses, ** significant at 5%, * significant at 10%. [2] Clustering by subject is appropriate because there are multiple observations for each subject due to the multiple-round experimental design. [3] Clustering by session does not change any of the significance levels in panels A and B. In panel C, clustering by session reduces the significance of the coefficient on good values in column 1 to the 10% level (p=0.061), and increases the p-values of the "good values" in columns 2 and 3 to p=0.198 and 0.125, respectively. [4] The numbers of observations falls in later rounds because some sessions, due to time constraints, contained fewer than 12 rounds. [5] The number of observations falls when including our "good values" measure, because some participants did not complete the survey. [6] *Initial own trustworthiness* is the average proportion of money received that a subject would return---averaged over each possible amount that could be received---measured the first time the subject was assigned the role of receiver. To minimize contamination of this measure of trustworthiness by learning, while still maintaining a reasonable number of observations, all regressions using this measure only include subjects who were an entrepreneur for the first time in one of the first two rounds. [7] *Good Values* is the average of two measures obtained from a survey that subjects completed either a week prior or a week after their experimental session occurred: i) the emphasis, on a scale from 0 to 10, that the subject's parents placed on being a model citizen as a value during their upbringing; and, ii) on the same scale, the emphasis their parents placed on group loyalty. We then divide the resulting average by 10 to put this measure on a scale comparable to beliefs (0 to 1). [8] *Expected Trustworthiness* is the average proportion each subject expected entrepreneurs to return within a particular round. Beliefs were elicited in an incentive-compatible manner for each possible investment level; the variable used is the average of these beliefs, for each subject, over each possible amount a receiver could receive. Beliefs were elicited regardless of the role the subject played in a particular round; if the subject was currently a receiver, they were instructed to exclude their own action from the calculation, and remunerated on this basis as well.

Table 3
Trust belief errors and economic performance in the on-line experiment
OLS estimates of sender's earnings on errors in trust beliefs

	(1)	(2)	(3)
Belief Errors	1.898	2.196	2.196**
	(1.595)	(1.577)	(0.742)
Belief Errors Squared	-24.061***	-23.360***	-23.360**
	(7.353)	(7.945)	(4.798)
Constant	11.465***	10.995***	10.995***
	(0.356)	(0.639)	(0.118)
Session Fixed Effects?	No	Yes	Yes
Session-Clustered Std Errors?	No	No	Yes
Observations	122	122	122
R-squared	0.05	0.07	0.07

Notes: [1] Robust standard errors are in parentheses, *** significant at 1%, ** significant at 5%, * significant at 10%. [2] Belief errors are defined by the difference between a participant's estimate of the proportion of money received that a receiver will return and the actual average return proportion within each session, averaged over each possible amount a receiver could receive. This value excludes the participant's own action in the role of receiver. This yields a number that ranges from -1 to 1 for each participant.

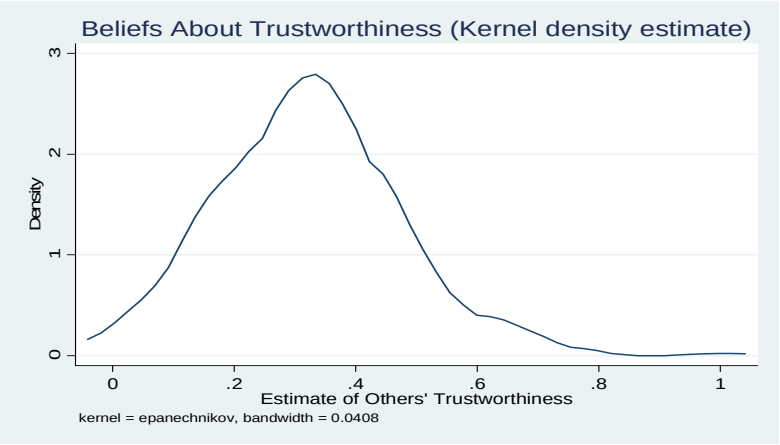
Table 4
Earnings by trust belief categories in the on-line experiment
OLS estimates of sender's earnings on dummies for trust beliefs categories

	(1)	(2)	(3)
Accurate Estimators	1.860***	1.773***	1.773**
	(0.663)	(0.657)	(0.500)
Over-estimator	0.311	0.324	0.324
	(0.706)	(0.681)	(0.352)
Constant	9.930***	9.554***	9.554***
	(0.525)	(0.603)	(0.135)
Session Fixed Effects?	No	Yes	Yes
Session-Clustered Std Errors?	No	No	Yes
Observations	122	122	122
R-squared	0.07	0.09	0.09

Notes: [1] Robust standard errors are in parentheses, *** significant at 1%, ** significant at 5%, * significant at 10%. [2] Dependent variable is sender's earnings in euros. [3] The excluded category is "under-estimators." [4] Belief error categories are defined as follows: "Accurate Estimators" had an average belief error within the interval [-0.1, 0.1]; "Over-estimators" had an average belief error in the interval (0.1,1]; "Under-estimators" had an average belief error in the interval [-1,-0.1). [5] We also considered wider and narrower intervals separating the three categories, using [-0.15, 0.15] and [-.05, 0.05] to define accurate estimators. This did not change anything qualitatively; [6] Another specification used the 33rd and 66th percentiles of the error distribution in the data to separate the three categories. This did not change the results.

Figure 1
Heterogeneity in trust beliefs and own trustworthiness

A. Trust beliefs



B. Own initial trustworthiness

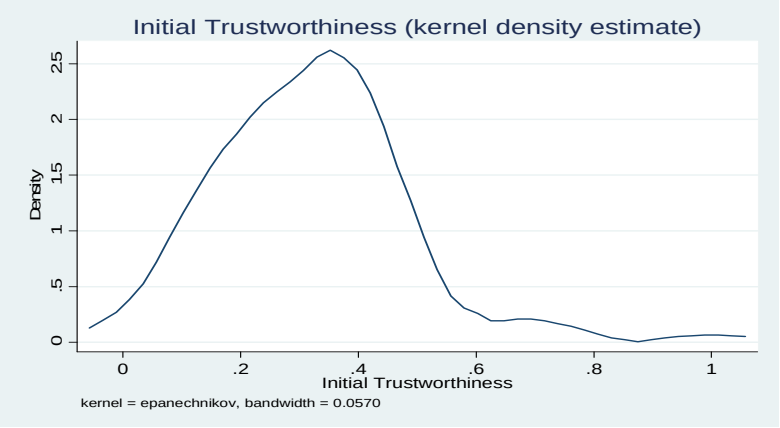
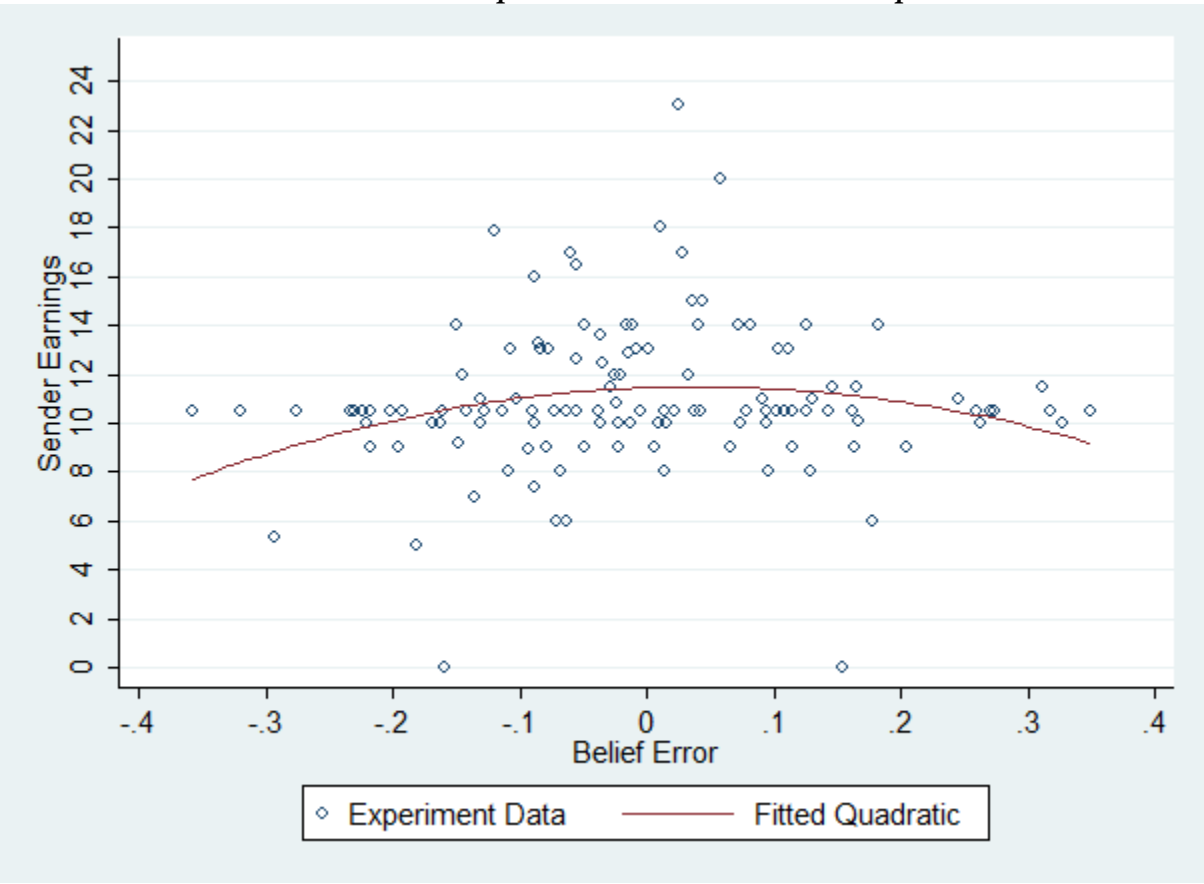


Figure 2
Trust belief errors and performance in the on-line experiment



Appendix: Trust Experiment Design

Laboratory Experiment

Participants were recruited from a pre-existing list of students who had previously expressed willingness to take part in experiments, in general, at LUISS Guido Carli University in Rome, Italy. All laboratory sessions were conducted at CESARE, the lab facility at LUISS. The experiment was programmed and implemented using the software z-Tree (Fischbacher, 2007).

After showing up to the lab at pre-scheduled session times, instructions were seated at individual desks in the lab, each separated by opaque dividers, and each equipped with its own computer. Instructions were then read aloud by the experimenters, and participants' questions, if any, were answered by the experimenters. This initial phase—instructions and seating—typically took from 15-30 minutes.

After questions were answered, subjects proceeded to the game-playing phase. This phase consisted of up to twelve rounds of the trust game, as described below. Participants were not informed how many rounds of game-play there would be, but rather only instructed that there would be "several" rounds. This was meant to mimimize end-game effects possible when the number of rounds is known. Because sessions were scheduled to last (up to) two hours, and because most participants had never participated in any experiment before (CESARE is a new facility) the number of rounds per session varied widely. Sessions consisted of anywhere from 3 to 12 rounds, with the majority consisting of 12 rounds.

Even though the experiment involved repeating the same game for multiple rounds, participants were randomly re-matched with an anonymous partner each round, and within each pairing roles were randomly reassigned. These design features allow for learning about the population's preferences but not about any specific person's preferences, as desired. It also ameliorates many repeated-game effects that are possible when partners are uniquely identifiable, or persist over rounds—such as reputation builiding or punishing/rewarding specific partners for past behavior—that, while important in the real world, are not the focus of this experiment.

The trust game is a two-player sequential-moves game of perfect information. The first-mover, called the "sender," is endowed with 10.50 euros. The second-mover—the "receiver"—has no endowment. The sender chooses to send some, all or none of his or her endowment to the receiver. Any amount sent is tripled by the experimenter before being given to the receiver. The receiver then chooses to return some, all or none of this tripled amount back to the sender, ending the game. Sending a positive amount entailed a small fee—0.50 euros.

Senders were allowed to send either 0 (euros), and retain 10.50, or send any positive whole-euro amount: $1, 2, \ldots, 10$. Receivers' decisions were collected using the strategy method. Before receivers discovered how much their sender sent, they specified how much they would return for any amount of money they could receive. Specifically, receivers were faced with a series of ten separate screens, each asking only one ques-

tion: "if you receive m euros, how much will you return?" For each separate screeen, m was replaced with exactly one value, $m \in \{3, \ldots, 30\} = \{3 \times 1, \ldots, 3 \times 10\}$. The order of possible amounts, m, was randomized in order to avoid inducing any artificial consistency in receivers' strategies and to make each decision feel as real as possible to receivers. This random order was the same for all receivers within each round, and was re-randomized between rounds. Obviously, no information about receivers' decisions was shared with senders in any way before the end of each round.

At the end of each round, each sender and receiver pair was informed of the outcome of their interaction only—i.e., how much the sender sent, and, if this was a positive amount, how much the receiver returned as determined by the relevant element of the receiver's strategy vector. No other elements of the receiver's strategy vector was revealed.

To collect beliefs, within each round every participant, regardless of the role they had been assigned, was asked to estimate the amounts receivers would return, on average, for each possible amount receivers could receive. Specifically, participants answered ten questions: "How much would receivers return, on average, if they were to receive m euros?", $m \in \{3, \ldots, 30\}$. Participants who were currently receivers were told to exclude their own actions from this estimate, and estimate how much *other* receivers would return, to rule out any mechanical—real or imagined—connection between own-actions and estimates.

Incentives to report beliefs truthfully were given by paying subjects according to a quadratic scoring rule. It is well-known that this rule gives (risk-neutral) individuals incentives compatible with reporting truthfully the mean of their subjective distribution of beliefs. Specifically, for each of the ten belief questions participants earned an amount of money given by the function below, where $\widehat{r_m}$ is receivers estimated return amount, r_m is receivers actual (average) return amount, and as above $m \in \{3, \ldots, 30\}$:

$$Earnings = 1 - (\frac{\widehat{r_m} - r_m}{m})^2$$

For example, if a subject's estimate of receivers' average return amount, conditional on receiving 9 euros, was 6 euros–i.e., $\widehat{r_9} = 6$—and receivers' strategy vectors entailed returning (on average) 2 euros conditional on receiving 9, then that participant's estimate would earn the participant (in euros)

$$1 - (\frac{6-2}{9})^2 = 1 - \frac{16}{81} \approx 0.80 \tag{1}$$

A perfect estimate paid 1 euro, so that subjects could earn up to 10 euros each round from their estimates. Beliefs were elicited either before or after participants submitted their actions, with this order being randomly re-determined for each participant before each round.

When all rounds were completed, one round was selected at random and participants were paid in accordance with their actions and the accuracy of their estimates

in that round. This procedure is meant to eliminate wealth effects from accumulated earnings over rounds and is standard in the literature. All of these design elements were (commonly) known by all participants.

On-line Experiment

Participants were recruited from the same pre-existing list of potential experimental student participants at LUISS in Rome, Italy. We excluded from the list of invitees anybody who had taken part in the laboratory experiment, so that no individual took part in both the in-lab and the on-line experiment. This experiment was conducted on four separate days, each day constituting a session. In total, 122 students participated in the on-line experiment.

The on-line experiment implemented one round of the trust game in the same manner as above with three exceptions. The first exception is that the function used to transform money sent into money received was no longer linear, but rather quadratic. This function was presented to participants in table form (below). Secondly, a full strategy method was used: participants submitted their decisions in both possible roles before learning which role they would be assigned. Finally, participants did not know their beliefs would be elicited until after they submitted their decisions. This weakens concerns that belief elicitation itself could affect decisions.

If the sender sends (euros):									
1	2	3	4	5	6	7	8	9	10
Then the receiver will receive (euros):									
8.05	11.30	13.85	16.05	17.90	19.60	21.20	22.65	24.05	25.30

In terms of earnings, two features are notable. First, belief accuracy was remunerated using a slightly different procedure: a randomized quadratic scoring scoring rule. Schlag and van der Weele (2009), among others, have proven that this procedure is theoretically robust to indvidual risk preferences. Specifically, each estimate is converted into a number, $z \in [0, 1]$, precisely as above. At the same time, the computer chooses at random a number, $y \in [0, 1]$. If $y \leq z$, the participant earns 5 euros, otherwise the estimate pays nothing. At the end of the session, one estimate is randomly chosen to count towards a participants potential earnings. The second feature of note is that only 10 percent of participant pairs were (randomly) chosen to be paid according to their decisions and estimates. Since the on-line experiment required much less of participants' time, this kept hourly earnings comparable to earnings in the laboratory experiment.

References

[1] Fischbacher, Urs (2007), "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*, 10(2), 171-178.

[2] Schlag, Karl and J. van der Weele (2009). "Eliciting Probabilities, Means, Medians, Variances and Covariances without assuming Risk Neutrality," mimeo