Unintended consequences of corruption indices

Chapkovski, Philipp

University of Bonn

15 March 2022

Unintended consequences of corruption indices

Philipp Chapkovski*

February 15, 2023

Abstract

Using the results of a pre-registered online experiment, this paper examines how information about a corruption in a group can affect intergroup relations. Corruption indices are not only a valuable tool for investors and policymakers to make informed decisions, but can also lead to statistical discrimination: Individuals from a more "corrupt" region may be perceived as less trustworthy or more prone to dishonest behavior. To test this hypothesis, we manipulated the amount of information participants had about their potential partners' regions of origin and asked them to (a) estimate the proportion of participants in each region who report a more profitable outcome in a coin toss game and (b) transfer money to a partner in each region in a trust game. The presence of a regional corruption index led participants to significantly overestimate the degree of dishonesty by participants from more corrupt regions and to reduce trust towards them. The results show how corruption indices can be a source of statistical discrimination against outgroups despite the well-meaning intentions of their creators.

Keywords: Corruption; Honesty; Trust game; Group identity; Beliefs; Russia
JEL classification: C91; C92; D63; D73; D84

*Philipp Chapkovski (chapkovski@uni-bonn.de) is a post-doctoral researcher at the Institute for Applied Microeconomics in the University of Bonn.

1 Introduction

What harm can come from the state providing citizens with more information about the country in which they live? After all, more transparency and more information will suppress corruption and improve the provision of public goods, such as education [Azfar and Nelson, 2007, Reinikka and Svensson, 2011]. But what if more knowledge leads to prejudice against our fellow citizens and does more harm than good?

In this paper, we focus on the potential negative consequences of a rather specific tool: a regional corruption index that ranks different regions based on the (perceived) corruption of their local bureaucracies. However, the mechanisms that can lead to statistical discrimination and distrust of entire groups may also be at work in other country-specific or regional indicators of governance quality.

When we think about the quality of governance, we usually have country-level indicators in mind, such as the World Bank's indicators [Mastruzzi et al., 2010]. Meanwhile, ranking regions by the level of corruption of their bureaucracies would be a valuable tool for many countries to provide individuals, businesses, and the state with the information they need to make more informed decisions.

Perhaps the Russian government was guided by these considerations when it adopted its National Anti-Corruption Plan in 2019. The document announced that the state would fund annual nationwide surveys to measure the perceived level of corruption in each of Russia's more than eighty regions [Butrin, 2019][1]. The goal was to create an index that could rank each Russian region along a spectrum of corruption. The index was based on a series of questions about the regularity of corrupt encounters and demands, the amount of bribes, and the willingness of respondents to pay such bribes, and was generally similar to an earlier regional corruption index once produced by the Russian Ministry of Economic Development[2]. Each of the 86 Russian regions was assigned a score from 0 to 100, with more points indicating a higher level of corruption. The index shows that corruption varies widely across regions, ranging from 15 points in the Jewish Autonomous Region to 80 points in the Krasnodar Region.

The logic driving the state to create such indices is clear: indices can be used as important performance indicators to measure the efficiency of regional authorities; this is particularly helpful for a highly centralized state like Russia, where local governors are essentially appointed from above. Various indices measuring corruption at the state level

[1] Probably due to the pandemic COVID -19 and the subsequent Russian aggression against Ukraine, this plan was put on hold and has still not been implemented by February 2023.

[2] The original data collected by FOM [2011] is available in Russian at: https://old.economy.gov.ru/minec/resources/116f09004739f0c7a2a4eeb4415291f1/doklad_kor.pdf

have been crucial in shedding light on widespread corruption, identifying its main causes, and paving the way out of it. Well-known indices such as Transparency International's Corruption Perception Index [Lambsdorff, 2007] and the World Bank Governance Indicators [Rohwer, 2009] have proven to be effective policy tools for guiding investors' decisions and pressuring both government and the public for reform [Christiane, 2006].

As useful as these indices may be, they come at a price. They can lead to a corruption trap: a vicious cycle in which international policymakers might freeze development aid to a corrupt country, making it more difficult to implement anti-corruption reforms due to reduced available funding [Andersson and Heywood, 2009]. Furthermore, in addition to direct potential harm, such indices limit a state's ability to improve its governance indirectly by suppressing foreign investment [Woo and Heo, 2009].

But the potential negative effects of corruption indices are not only felt at the macro level. The very existence of such indices can change the environment they are designed to objectively measure. They can affect the everyday decisions of citizens, creating a kind of Hawthorne effect [McCarney et al., 2007], i.e., changing one's behavior when one knows that one is being watched.

For instance, if the information that society as a whole is mired in corruption becomes widely known, bribery becomes a normatively permissible behavior. Data from the European Values Survey showed that people are more tolerant of corruption if they believe that their social environment is involved in corrupt practices [Dong et al., 2012]. This finding was confirmed in a survey experiment in Costa Rica, which showed that people were more willing to bribe a police officer if they knew that more and more of their fellow citizens admitted to having faced corruption in the past 12 months [Corbacho et al., 2016]. Knowledge of high levels of corruption can also undermine the public's will to fight corrupt bureaucracies; an analysis of data from 71 countries in Transparency International's Global Corruption Barometer survey showed that the perception of high levels of corruption can demotivate citizens to actively resist it [Peiffer and Alvarez, 2016].

Many people refuse to pay a bribe because they have an intrinsic moral stance against it. But as is the case for social norms in general, this anti-corruption norm is susceptible to social pressure: it is shaped by a person's beliefs on how many others would pay a bribe in a similar situation (empirical expectations) and what they think others believe is the right thing to do (normative expectations) [Bicchieri, 2005]. When they learn that the society they live in is more corrupt than they think, both normative and empirical expectations change, lowering the moral cost of such an action [Cheeseman and Peiffer, 2020].

In this paper we focus on corruption indices as a possible cause of discrimination between groups. If we know that a person we are dealing with comes from a high corruption area, we may trust him less and suspect that he is prone to morally dubious activities. In other words, corruption indices can trigger statistical discrimination [Fang and Moro, 2011], creating and reinforcing prejudices against people from certain regions.

Most studies addressing the shortcomings of existing indices use country-level measurements [Donchev and Ujhelyi, 2014, Warren and Laufer, 2009]. In this paper, we intentionally choose a relatively rare regional corruption index as the main manipulation tool. In doing so, we start from the premise that regional stereotyping is much lower than cross-country stereotyping. Thus, prior to manipulation, preexisting sentiments toward different regions would be relatively weak or nonexistent. By using only regional differences within a country, we can ensure that the country factor remains fixed. If those who trust are from the same country as those to be trusted, we can control for any cultural biases toward different countries, including pre-existing beliefs about levels of corruption, honesty, and trustworthiness. Moreover, this work can contribute to the otherwise understudied subfield of variation and perceptions of regional corruption, which has rarely been the focus of research due to a lack of data (the work of Charron et al. [2015] is a notable exception).

This paper brings together two strands of literature: one that analyses the unintended consequences of information about corruption, and another, more extensive set of studies that examines the factors that influence human honesty, propensity to deceive, and trust. Specifically, we want to examine how knowledge about the regional extent of corruption leads people to reevaluate their beliefs about the honesty and trustworthiness of the inhabitants of a given region.

Much like the intention to pay a bribe, the propensity to cheat – or conversely, to be honest – is a social norm that depends on what others do (or what people think they would do). In other words, a person's propensity to lie is determined to some extent by his/her estimated proportion of others who lie under similar circumstances. Despite the extensive literature on this topic, results remain inconclusive as to the direction and magnitude of the effect. Some studies have found a positive effect (people cheat more when they observe others cheating) [Diekmann et al., 2015, Necker et al., 2020]; others have found that the link between beliefs and cheating is complicated: Beliefs about cheating *per se* do not make people more inclined to cheat, but exposure to actual cheating pushes people to adjust their behavior to bring their behavior in line with that of the majority [Rauhut, 2013].

To the best of our knowledge, this is the first study to examine how information about the prevalence of corruption in a group changes beliefs about how honest and trustworthy those group members are. But the relationship between corruption and honesty has been studied before. Country-level corruption scores correlate significantly with beliefs about dishonesty. People expected higher levels of dishonesty in more corrupt countries, although their actual behavior did not correlate with their own country's level of corruption [Mann et al., 2016]. In another study, individual cheating behavior was observed more frequently in countries ranked as more corrupt on the Corruption Perception Index [Jiang, 2014]. A cross-national study involving officials in 10 different countries also found that country-level corruption indicators were strongly correlated with average behavioral dishonesty [Olsen et al., 2019].

In this study, we analyzed decisions of a group of participants from Moscow, matching them with their counterparts from three Russian regions scoring from low to high in the regional corruption index. We manipulated the amount of information available about each region, and we tested whether the availability of regional corruption index changed individuals' estimates of other participants who lived in those regions. We found out that provision of regional corruption level encourages people to significantly raise their estimates of other people's dishonesty. We were also able to track the effect of corruption information on trust toward partners from these regions. Information about corruption decreased trust towards their partners. Both effects on dishonesty estimates and trust were even more pronounced when participants were able to choose themselves what kind of information they want to observe about each region.

We begin with some theoretical considerations on which we build our pre-registered hypotheses (section 2). In section 3 we describe the experimental setup and the data collection procedure. In section 4, we present the results and show how they match our pre-registered hypotheses. We conclude with section 5, where we discuss the limitations of this study and the possible further avenues of research to which it leads.

2 Theoretical considerations and hypotheses

Before running the study, we pre-registered four hypotheses[3]:

Effect of corruption information on outgroup estimates of honesty and trust

The causal relationships between corruption, trust, and honesty are intertwined. Because corrupt transactions are by definition unenforceable contracts, they require a degree of

[3]The preregistration is available at the AsPredicted web server at https://aspredicted.org/a923e.pdf)

trust between the parties. Therefore, scholars assume that a certain type of trust (e.g., 'particularistic' trust, as between conspirators in a corrupt deal) promotes corruption [Uslaner, 2004] and mitigates its negative effects on economic growth [Li and Wu, 2010].

On the other hand, corruption erodes trust: Those who were asked for a bribe in the corruption game had lower beliefs about how trustworthy other people are in the subsequent trust game [Banerjee, 2016]. Similarly, a personal experience of dishonesty impairs trust. People who were lied to (in a Deception game) were less likely to trust in a subsequent trust game, even when matched with others who had not lied to them [Gawn and Innes, 2018]. Dishonesty also weakens trust in sanctioning institutions: In trust and dictatorship games with a third-party punishment, people showed less trust and altruism when they knew that a punisher had behaved dishonestly [Spadaro et al., 2022]. All of these causal links between dishonest or corrupt behavior and trust presuppose an individual experience with these phenomena. This paper conjectures that even more general information, such as about the quality of governance in a region, forms expectations about potential dishonesty and trustworthiness of individual representatives of that region.

The regional information can be important because identities based on geographical or national boundaries may create groups with strong tendencies for ingroup bias. In a Trust game played between representatives of different districts of Zurich, participants showed a clear preference for their own group by trusting more participants from their own district [Falk and Zehnder, 2013]. Expectations of honesty, which are critical to building trust, are also based substantially on group identities such as country of origin. Unlike trusting behavior there is a surprising negative ingroup bias: People expect other people be less honest if they are culturally close to them [Dieckmann et al., 2016]. Comparison of beliefs regarding honesty in fifteen countries also demonstrated that "these beliefs are driven by biases, including self-projection and, surprisingly, pessimism about the honesty of people in one's own country" [Hugh-Jones, 2016]. We hypothesise that group identity based on geographical boundaries interacts with additional information regarding the corruption level in each region. That may exacerbate the degree of outgroup discrimination towards more corrupt regions.

H1 *With additional information about the regional level of corruption, people provide lower honesty estimates about people from more corrupt regions.*

H2 *With additional information about the regional level of corruption, first movers in a trust game will make lower transfers toward the second movers from more corrupt regions.*

Effect of corruption information on image of one's own group

Information about how widespread corruption is, also changes people's expectations about their own group morality. It causes people to reevaluate the risk of engaging in bribery. The effect of information on corruption level is asymmetric: There is contagion (people offer more bribes when observing others bribing), but not conformism (observing others bribing less does not drive people to offer less bribes) [Schram et al., 2022]. The possibility of manipulating willingness to engage in corrupt practices by influencing beliefs about those practices is tempting, although the effectiveness of such nudging is unclear. On the one hand, reducing beliefs about the extent of corruption makes participants less willing to engage in corrupt transactions [Köbis et al., 2015, 2022]. On the other hand, these results might be clouded by the recent finding that information about the honesty level of others does not have a positive effect on one's own honesty [Dimant et al., 2020]. Beliefs about honesty of others in one's own group is an important factor that drives personal honesty: "If a majority of one's peers are perceived to be honest, an individual is likely to suffer a larger aversion penalty/disutility when behaving dishonestly" [Robert and Arnab, 2013]. This does not work in the opposite direction: Personal experience with cheating does not always affect one's assessment of how much others would cheat. A field experiment with fare dodgers in Italy showed that whether a person had evaded fare or not did not affect his or her beliefs about the proportion of fare evaders [Bucciol et al., 2013]. In this study, we expect to observe the effect of overestimating dishonesty within one's group when participants are exposed to the level of corruption in their home region. This would provide a link between studies of the effects of beliefs on corruption and dishonesty and reveal a potential spillover effect between these two domains.

H3 *Participants' estimates of the honesty of people from their own region are different when we give them the information about the corruption level of their own region.*

H4 *The transfers of first movers in a trust game toward people from their own region are different when they are additionally provided with information about the corruption level of their own region.*

Selection of information

The presence of the regional corruption index may influence respondents' assessments not only directly, but also through the experimenter's demand [Zizzo, 2010]. Respondents might try to guess the intentions of the researchers and act accordingly. Although in our experiment we disguised our research objectives by blending the corruption information with other information about the region, there was still a risk that the outcome was influenced by the experimenter's demand. To overcome this inherent design limitation, we allowed participants to self-select the regional information in one of the treatments.

We did not pre-register any hypotheses about the effect of endogenous selection on trust and honesty beliefs because we expected that this would rule out a possible demand effect and allow us to test the above hypotheses more cleanly. However, based on previous findings the ability to select information almost always increases the discriminative effect. When people have the opportunity to choose their own partners before performing a task with real effort and the possibility of misreporting results, participants who have cheated individually choose dishonest partners and use their presence to further justify their dishonest behaviour [Charroin et al., 2022]. The study of gender discrimination in a trust game found no discriminatory effect in treatments where participants were randomly matched, but strong discrimination when they could choose their own partners [Slonim and Guillen, 2010]. Finally, the ability to choose which group-level honesty characteristic to use (average vs. maximum), significantly increased individual dishonesty [Akın, 2019].

3 Experimental design and procedure

3.1 Experimental design

Our main objective in this study was to measure how the presence of information about the general level of corruption in a group may alter judgments of the honesty and trustworthiness of those group members. Depending on the treatment, participants either had access to information about the extent of corruption in the region or they did not. We measure estimated levels of honesty by eliciting beliefs about the estimated proportion of participants from each region who would indicate heads after flipping a coin (an outcome that would increase their personal payoff). The level of trust is captured by a transfer made by a first mover (trustor) in a Trust game [Berg et al., 1995]. In all three treatments, respondents participated in both the belief elicitation and the trust game (in randomized order).

Two-pools approach

To keep fixed an in-group membership of a group whose honesty beliefs and trust we wish to measure, we divided the total population recruited for participation into two pools: a *source* pool and a *target* pool. The source pool was the group whose preferences and

Region	N	Treatment	Information shown to participants
Moscow	200	FIN	GRP, Age, CPI
Moscow	200	FIC	GRP, Age, CPI, **Corruption**
Moscow	200	EI	*3 out of 6 could be chosen*: GRP, Age, CPI, Corruption, Unemployment, Birth rate

(a) *Source* pool characteristics

			Regional information					
Region	N	Treatment	GRP	Age	CPI	Corruption	Unemployment	Birth rate
Moscow	100		1,423,589	41.89	103.4	58	1.4	10.7
Arkhangelsk	100	NA	712,653	40.15	103.3	17	6.3	9.0
Voronezh	100		404,839	41.92	102.6	81	3.6	8.4

(b) *Target* pool characteristics

Table 1: *Source* and *Target* pools information. *GRP* - Gross regional product per capita; *CPI* - Consumer Price Index; *Age* - an average age of inhabitants in a region; *Corruption* - Perceived Corruption Index [FOM, 2011]; *Unemployment* - Unemployment rate in %; *Birth rate* - Number of newborns per 1.000 women. With an exception of Corruption Index, all the data were taken from Federal State Statistics Service regional database for 2020. An order the items were presented were randomized across participants. *FIN*- Fixed neutral information treatment; *FIC*- Fixed information with corruption treatment; *EI*- Endogeneous choice of information treatment.

beliefs we intended to measure. It was recruited exclusively from the Moscow region and randomly assigned to one of three treatments with different exposure to information about corruption. The target pool was residents of three different regions (including Moscow) with different positions in the regional corruption index. After participants in a source pool made their choices and expressed their beliefs about participants in each of the three regions of the target population, we matched their choices and beliefs with the actual choices of the target pool and calculated the payoffs for both pools. This two-step procedure allowed us to avoid deception and make the choices for two pools independent of each other in time.

The decisions of the target pool of each region were used only to calculate the payoffs and are not considered in the rest of the analysis: Only the source pool decisions were used to test our hypotheses.

Treatments

Participants from the *source* pool were randomly assigned to one of three treatments: Fixed Neutral Information (*FIN*), Fixed Information + Corruption (*FIC*), and Endogenous Choice of Information (*EI*). The treatments differed only in the amount of information provided about each region from which the *target* pool was recruited. In fixed-information treatments (*FIN* or *FIC*), participants received a predefined amount of information about each region. In the endogenous choice treatment (*EI*), participants could select the indicators they wanted to see. The order in which each piece of information was shown was random for each participant. The order in which the available items were shown for further selection was also randomized. The composition of the source and target pools and the information available in each treatment are shown in Table 1. Examples of screenshots with provided information for each region are available in Appendix (Figure A.2).

Measurements of beliefs about honesty

We operationalize beliefs about honesty by eliciting participants' estimates of the percentage of others who would report that they had observed heads after flipping a coin [Bucciol and Piovesan, 2011]. The belief elicitation was incentivized: each correct guess by a participant increased the payment that they received [Krupka and Weber, 2013].

Measurements of trust

We measured trust using participants' transfers as first movers in a trust game. In a standard trust or investment game [Berg et al., 1995], participants are matched in pairs. Both players are given a specified initial amount (endowment) and are assigned one of two asymmetric roles. The first player (*trustor*) can choose to give the second player (*trustee*) any fraction of the endowment. This amount is then multiplied by the experimenter by a certain coefficient $k > 1$ (in our case $k = 3$). The second mover can send back any amount from the received multiplied sum. Thus, the amount a first mover sends reflects its belief that the second mover will at least partially reciprocate its trust, while the amount the second mover sends back reflects his/her actual trustworthiness. In a purely profit-maximising strategy, a trustee would send back nothing of the multiplied amount received from the trustor. Since there is an expectation that this trust will not be reciprocated, a rational trustor transfers zero to the trustee, but worldwide, trustors send an average of 50.2% of their endowment, while trustees return 37.2% [Johnson and Mislin, 2011].

Regional Corruption Index and choice of specific regions

We selected three regions to recruit our target pool and thus to capture the beliefs of a source pool: Arkhangelsk, Moscow, and Voronezh. Correspondingly they are ranked at the top, middle and bottom of the index of domestic corruption. The corruption index varies from 0 (the lowest possible level of perceived domestic corruption) to 100 (the highest degree of corruption). Index values for all three participating regions are shown in The table 1b. The index used in this study is based on the 2011 report by the Ministry of Economic Development [FOM, 2011]. The data used to compose the index of regional corruption were collected by the Public Opinion Foundation, one of Russia's largest pollsters. The survey was conducted in 70 Russian regions ($n = 17.500$), representing 94.5% of the population. The index was formed by summing the frequency of responses in each region along four dimensions: (1) the risk of being asked for bribes, (2) the proportion of those who have ever paid bribes, (3) the willingness to pay bribes, and (4) the average amount of bribes paid. It should be noted that the composition of the index and its value for each region are not in themselves relevant to the purpose of this study: Our goal was to determine whether knowledge about each region's performance according to this index influences behavior, not the index itself. What is crucial, however, is the fact that Russia in general has a high degree of heterogeneity in perceived levels of corruption across regions (see map of Russian regions in the Appendix, Figure A5), where the Arkhangelsk region has an index value four times lower than that of the more corrupt Voronezh region.

To test whether corruption-related information about one's region can change estimates of honesty and trustworthy behavior, we also included the Moscow region, which was a home region for the *source* pool. How a stranger from a particular region is perceived may be influenced not only by information about the region in general, but also by the region of origin of the person making the assessment and by personal connections to the valued region. Since the only difference between treatments is the presence or absence of a regional corruption index, the difference in honesty ratings and trusting behavior toward one's region among Moscow residents could only be explained by this manipulation (see hypotheses H3 and H4). In addition, we control for a person's familiarity with each of the three regions with a series of questions listed in the Appendix (see figure A3).

3.2 Stages of the study

The study consisted of four distinct phases: the coin flipping game, providing information about regions, eliciting beliefs about others' decisions in a coin flipping phase, and the trust game. For a source pool, the order of the coin flipping and the trust game was randomized: Half of the population first reported the outcome of the coin flipping, then received information about regions and shared their beliefs about coin-flipping game across three regions; then they made their first-mover decisions in a trust game. The other half first received the information about each region, then made a decision in a trust game; after that they proceeded to the coin-flipping stage followed by belief elicitation. In Figure A4 (see Appendix), the flowchart shows the screen order for the two randomization outcomes [4].

Participants in a *target* pool reported only what they observed after flipping a coin, and then made a decision about what proportion of an amount received they would return to a first mover in a trust game. For participants in the *target* pool, the coin flipping and trust game phases were not randomized.

Below is a detailed description of each stage, with the coin flipping stage coming first.

Cheating (or coin-flipping) game, *CG* : In the standard coin-flipping game (based on Bucciol and Piovesan [2011]), participants were asked to flip a coin and report the results. For those who did not have a coin at hand, we provided a link to a search engine to find a virtual flipping coin from one of the numerous online services available. If participants reported heads, their payoff was increased by US$1; if tails were reported, the payoff remained unchanged.

Information about regions revealed, *RI* : After tossing the coin, participants received a three-column table with indicators describing each region. The indicators varied by treatment: the list of indicators available in each treatment is shown in Table 1a, and the specific values for each region are shown in Table 1b. In the Endogenous Choice of Information treatment, this information screen was preceded by a page where participants could choose 3 out of 6 available pieces of information. The order in which the regions were shown and the order in which the indicators were listed for each region were randomized for each participant but kept fixed across all screens for each participant. The same information was shown to participants in both the belief elicitation phase of the coin flipping game and during the transfer decision in the trust game. Screenshots of these stages can be found in the Appendix (Figures A1 and A2).

[4]The complete original instructions (in Russian) and the English version (translated by DeepL translation service) are available as supplementary online material. The full code for an experiment performed using the oTree platform [Chen et al., 2016] is available at `https://github.com/chapkovski/rc`

Belief elicitation about coin-flipping stage *BE* : Here we asked participants for their opinion on the proportion of others reporting heads. For each of the three regions, participants provided their estimate of how many out of 100 *target* pool participants would report heads. For each correct estimate (within a 10 percentage point margin of error), they could receive an additional 1$ bonus. Participants could see the information for each region as they made their estimates. (See Figure A1 for an example of this stage in the fixed information + corruption (*FIC*) treatment.)

Trust game - First move, *TG* : After participants completed the belief elicitation phase, they were informed that the second part of the study would begin. The instructions for the standard trust game [Berg et al., 1995] were shown, and participants had to pass the comprehension test. They were then informed of their role. Participants in the *source* pool were given the first-mover role (neutrally referred to as *participant A* in the instructions), and participants in the *target* pool were given the second-mover role (*participant B*). First movers were told that their decisions would be matched against the decision of a participant who might be from one of the three regions; thus, they had to make the transfer decision for each region, but only one would be relevant. First movers could see the same information for each region that they could see before and during the belief elicitation stage, in the same order of indicators and regions. Figure A2 shows an example of the first mover decision stage for the fixed information and corruption (*FIC*) treatment.

Regardless of whether the coin flipping or the trust game occurred first, participants were unaware of the content of the second part of the study at the beginning of the first part. However, they were informed at the beginning of the study that they could increase their final bonus in either part, but that only one of the two parts would be randomised and paid. This randomization of payments was done to avoid a hedging and wealth effect [Charness et al., 2016]. While making their decisions about reporting the results of the coin flipping, participants were not aware of the upcoming phase in which they would have to express their beliefs about what others would report.

Matching and payoff calculations

In the coin flipping part, we asked participants from the source pool to estimate how many participants from the target pool in each region would report heads after flipping a coin. To calculate payoffs for the belief elicitation, we matched the source pool's beliefs to the target pool's actual decisions in each region after the target pool reported the outcomes of coin flipping.

Similarly, in the trust game, first movers (trustors) could receive their payments only after second movers (trustees) had made their decisions to return transfers. For this reason, participants in a source pool made their decisions using the strategy method [Brandts and Charness, 2011]. First movers were informed that they could be matched with a second mover (*trustee*) who could be from one of the three regions, so they had to make the transfer decision three times; however, only one of these decisions would be implemented based on their partner's region. Second movers (trustees) from the target pool were asked to decide what share (from 0% to 100%) of the transferred and multiplied amount they wanted to send back to a first mover. After collecting the decisions from both pools, we matched the transfers from first and second movers and calculated the corresponding payouts. Because fewer pools are available in the Arkhangelsk and Voronezh regions, we matched the decisions of two first movers from the *source* pool with the decision of one second mover from one *target* pool. Thus, the decision to return a certain share of the multiplied amount to a first mover affected two random participants in a source pool, but only one (randomly chosen) first-mover transfer determined the payoff of a second mover.

The experimental design was evaluated and approved by approved by German association for Experimental Economic Research [5].

3.3 Procedure

Source pool sessions were conducted on November 17 (*FIC* and FIN treatments) and November 18, 2021 (*EI* treatment). Participants for the experiment were recruited via the crowdsourcing platform Toloka [Chapkovski, 2023]. Data were collected via the oTree server [Chen et al., 2016][6] . There were a total of 599 participants in a *source* pool who fully completed the study (286 (47.7%) of whom were women). Participants were randomly assigned to one of three treatments (FIC, FIN, EI). 200 participated in the FIN treatment (women: 98, 49%), 200 in the FIC treatment (women: 84, 42%), and 199 in the EI treatment (women: 104, 52%). Treatments were balanced with respect to socioeconomic characteristics: joint $F-$ and $\tilde{\chi}^2$ test results showed no differences (see details and test results in Table A1 in the Appendix).

Sessions lasted an average of 19 minutes (SD: 9.88). The show-up fee was $1 and was paid immediately; the variable part of the compensation (bonus) was paid after participants submitted their decisions and averaged $1.52 (SD 0.844). One participant did not submit the final completion code (although he/she completed the study) and therefore

[5] GFEW certificate number `bwcw68Gx`, available at https://gfew.de/ethik/bwcw68Gx

[6] The original data, the R code to generate the graphs and tables in the 'Results' section, and the instructions are available in the OSF repository: https://osf.io/su3dr/)

was not eligible for the bonus. Of the 598 participants, 304 were randomly selected for payment for the coin flipping part and received \$1.71 as a bonus (SD: 1.04), and 294 were randomly selected for the trust game part and received an average bonus of \$1.32 (SD: 0.49).

Target pool sessions were held on November 17 (Arkhangelsk and Voronezh) and November 18, 2021 (Moscow). Of the 300 participants invited, 296 took part in the study (138 women): 97 from Arkhangelsk (women: 37), 101 from Voronezh (women: 54), and 98 from Moscow (women: 47). The balance tests are not included here because the target group was not assigned to any treatment and their decisions were not included in the analysis.

The target pool sessions were shorter (as they did not make strategy decisions in a Trust game and did not participate in the belief elicitation stage of coin-flipping part) and lasted an average of 14 minutes (SD: 7.27). They received the same show-up fee of \$1 as the source pool. Their average variable earnings (bonus) were \$1.54 (SD: 1.00). Participants from Moscow received an average bonus of \$1.59 (SD: 0.95), Voronezh: \$1.59 (SD: 1.04) and Arkhangelsk: \$1.44 (SD: 1.00).

4 Results

Individuals beliefs in the honesty of others and trust, measured as transfer of first movers in a trust game, failed all tests for normal distribution. Shapiro-Wilk tests and Q-Q, distribution, and density plots for each dependent variable confirming this can be found in the Appendix (A.5 section). Therefore, nonparametric tests are used below to test for differences in means between treatments. For between-subject tests we used Mann-Whitney-Wilcoxon test (MWW), for within-subject comparisons we used paired samples Wilcoxon Rank Sum (WRS) tests.

H1 *With additional information about the regional level of corruption, people provide lower honesty estimates about people from more corrupt regions.*

We calculated for each participant the *belief_diff* variable, defined as the difference between her beliefs about overall honesty, that is, the difference between estimated number of heads reported per 100 participants in Voronezh (a reportedly highly corrupt city) and Arkhangelsk (a city ranked low on the corruption index).

On the left side of Figure 1, we show the difference in honesty beliefs: When corruption information was available, participants believed, on average, that the proportion of those who reported heads after flipping a coin would be 5.02 points

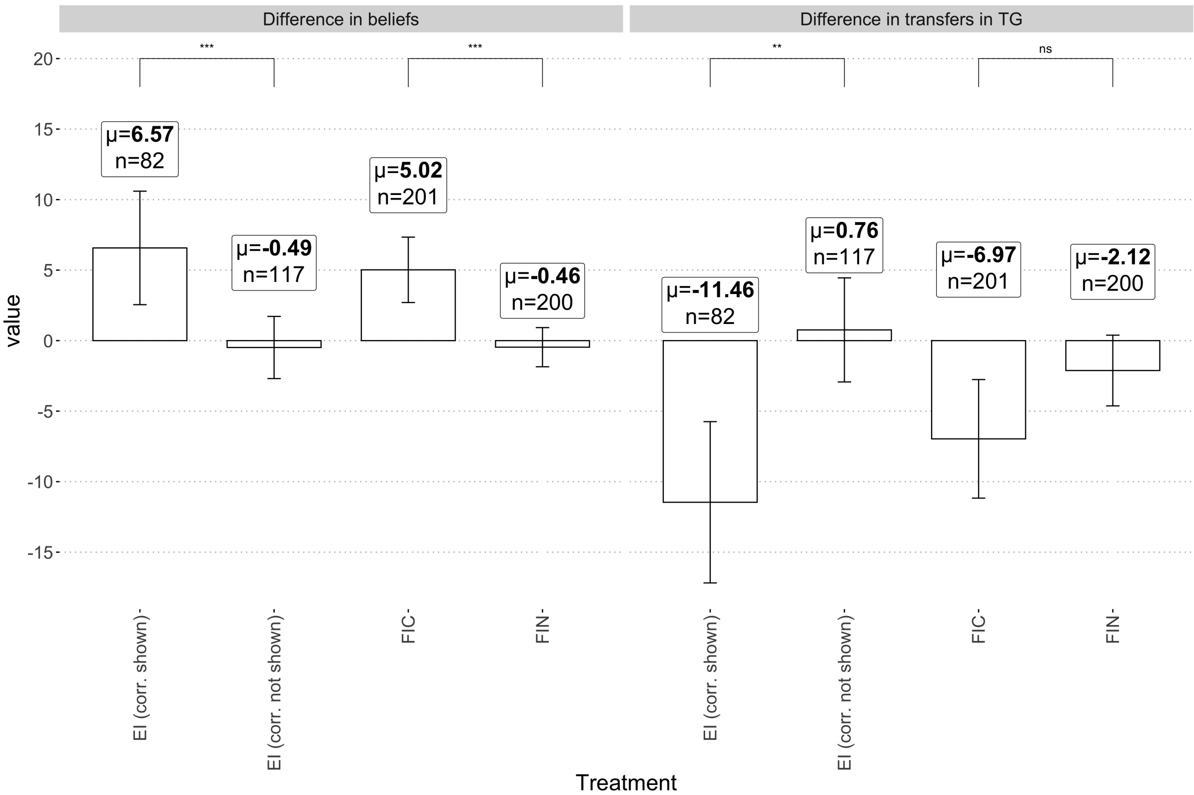


Figure 1: **Difference in beliefs and trust**. Difference in beliefs is calculated as a difference between individual estimates of how many will report head in a high-corrupt city (Voronezh) and a low-corrupt city (Arkhangelsk) within a single subject. Similarly, difference in trust is calculated as a difference between first mover's (trustor's) transfers to a potential partner in a high-corrupt city (Voronezh) and a low-corrupt city (Arkhangelsk). *TG* is a Trust game. The whiskers show 95% confidence intervals, labels show the means (μ) and number of observations (n). Mean comparisons show results of Mann-Whitney tests and symbols indicate statistical significance: ns: p > 0.05, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001. *FIC*– treatment with a fixed information set including regional corruption. *FIN*– treatment with a fixed information set without regional corruption. *EI*– endogenous information treatment where participants were able to choose information themselves.

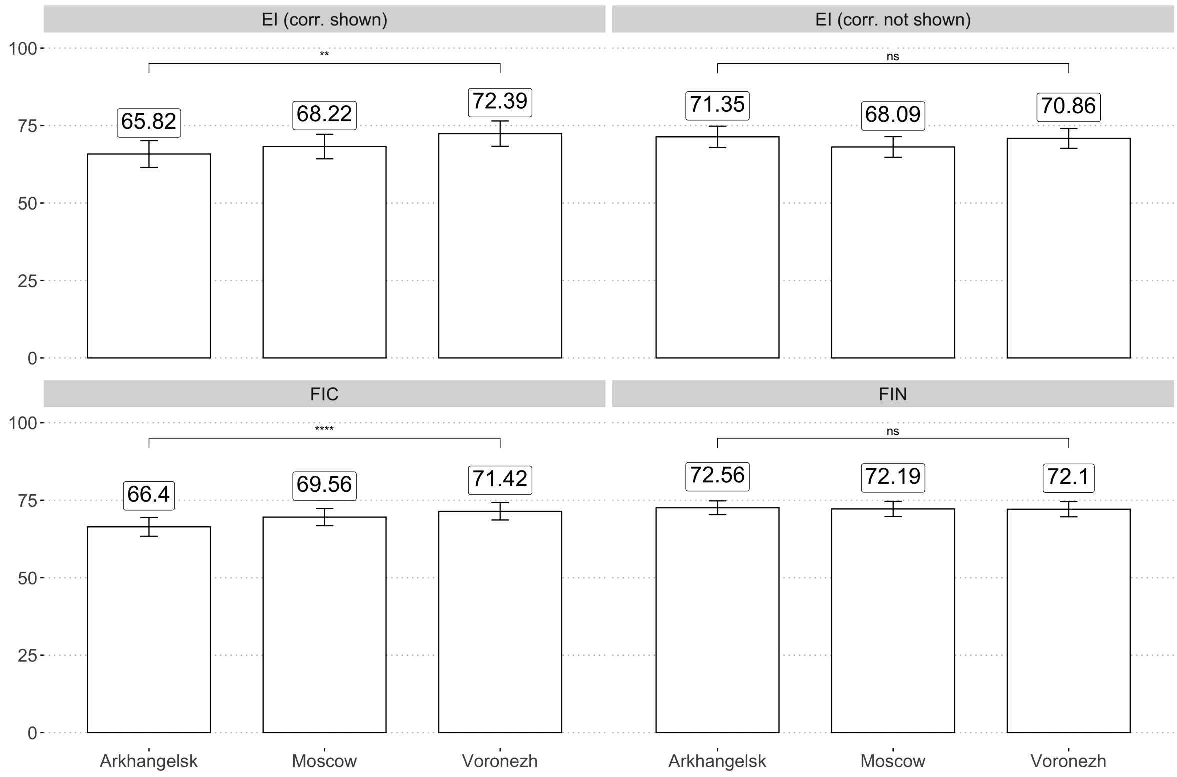


Figure 2: **Average beliefs about cheating game outcomes**. Each participant was asked to estimate the share (from 0 to 100) of reported heads in a coin-flipping (*cheating*) game among the participants living in on of three cities (Arkhangelsk, Moscow, Voronezh). The whiskers show 95% confidence intervals, labels show the means. Mean comparisons between Arkhangelsk and Voronezh show results of paired samples Wilcoxon Rank Sum tests and symbols indicate statistical significance: ns: p > 0.05, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001. *FIC*– treatment with a fixed information set including regional corruption. *FIN*– treatment with a fixed information set without regional corruption. *EI*– endogenous information treatment where participants were able to choose information themselves.

higher in a more corrupt region. This gap is significantly larger (MWW: $p = 0.002$) than the one in *FIN* treatment where regional corruption information was not available: there discrepancy between beliefs about Voronezh and Arkhangelsk -0.46 points.

This gap widened in the *EI* treatment, where we compared the beliefs of those who chose to observe the corruption index with those who did not: The belief gap for the former is 6.57 points and for the latter -0.49 (MWW: $p = 0.002$).

Figure 2 shows individual beliefs about each target region across treatments. In the *FIC* treatment with the corruption information participants expect that 66% will report head in Arkhangelsk, against 71% in Voronezh (WRS: $p < 0.0001$), a

7.5% growth. Meanwhile in a *FIN* treatment with no information this difference was negative 72.56 in Arkhangelsk vs. 72.1 in Voronezh (WRS: $p = 0.554$).

Again, the effect of corruption information became even stronger in the *EI* treatment, where participants could choose for themselves what type of regional information they wanted to observe. There those who chose to see the corruption information expected that 66% will report head in Arkhangelsk, against 72% in Voronezh (WRS: $p = 0.00046$), a 10% growth, and those who chose no to see believed that on average 71.35 people would report head vs. 70.86 in Voronezh (WRS: $p = 0.455$) (see detailed results of WRS tests in Appendix, table A5).

To test H1, we compare beliefs about the honesty of two regions made simultaneously by each person. The naive approach would be to analyse the constructed variable, *belief_diff* using an Ordinary Least Square (OLS) model. It is problematic though, because of the heavy problem of zero inflation: a substantial part of the population does not discrimate across two regions (Voronezh, and Arkhangelsk) so there is a zero difference between two estimates. We do report however the results of OLS models with different numbers of control variables in the Appendix (table A3).

Because most of the assumptions required for OLS models are not met and, in addition, each participant simultaneously estimates the proportion of heads reported in each region, this type of nested structure necessitates the use of generalized linear mixed model (or GLMM) for the analysis. In this model, the assumption of homogeneity and independence of sampling units is removed [Sciandra and Spera, 2022] - for the detailed analysis of non-normality of the data see Appendix, Section A.5.

An additional challenge in analysing honesty beliefs is that people report their believed proportions, which by definition are both downward and upward bounded. Based on suggestions by Ferrari and Cribari-Neto [2004] for analysing bounded shares and skewed distributions, we use a regression model based on the Beta distribution. Beta distribution is a continuous probability distribution defined in the range $f \in (0, 1)$, whose density function is defined as:

$$f(y, \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, 0 < y < 1 \tag{1}$$

with a mean, $E(f) = \mu$. The limits imposed by the range of bet function dictate the necessity of the following transformation of a dependent variable:

$$B^* = (B + 1)/102 \tag{2}$$

where B^* is the transformed ($0 < B^* < 1$) and B is the untransformed dependent variable (varying from 0 to 100).

We add random effects to the beta regression model to get the Beta GLMM [Bonat et al., 2015]. We use the standard link logit function, that connects the set of explanatory and control variables to the conditional mean, μ in 1:

$$\mu = y_{i,r} = \alpha + \beta_R r_i + \beta_T Treat + \beta_{RxT} r \times Treat + \mathcal{B}\mathbf{X}_i + \mu_i + \epsilon_{i,r} \tag{3}$$

where $y_{i,r}$ is an i's individual belief about share in a region r (transformed according to 2), μ_i is a random component for each participant, and $\mathcal{B}\mathbf{X}_i$ is a vector of individual factors (depending on the model specifications).

Table 2 reports the results of beta GLMM regression models in which the dependent variable is the belief in the proportion of respondents reporting heads in the coin flipping stage. The baseline target city here is Arkhangelsk, and the baseline subtreatment is the non-information treatment, *FIN*. The three models shown in Table 2 differ only in the degree of control. Model 1 is the simplest model, testing for interactions between target city and treatment without additional controls. Model 2 additionally controls for age, gender, the order in which the trust and coin-flipping stages were played, and participants' self-reported coin toss scores. Finally, Model 3 also controls for education level, marital status and employment status, income level, and the difference in knowledge about the two regions [7].

In all three models, beliefs about honesty in Voronezh in with-information treatments (*EI-corr* and *FIC*) are substantially higher than in the no-information treatments. The order of the game did not play any substantial role. The only factor that was significant apart from treatments was the participants' own decisions in the *CG* stage: those who reported heads have significantly higher odds of providing higher estimates of others who report heads. That makes our finding similar to those of [Mouminoux and Rullière, 2021] who also found that people who cheat tend to have higher beliefs about the dishonesty of others.

H2 *With additional information about the regional level of corruption, first movers in a trust game will make lower transfers toward second movers from more corrupt regions.*

[7] Detailed descriptions of each variable, including their internal names (used in the R code to create the regression tables and graphs in the text), labels and distributions for categorical variables, and distribution histograms for numeric variables are available in a codebook in the online supplementary materials.

	Model 1	Model 2	Model 3
(Intercept)	1.073***	0.632***	0.540**
	(0.073)	(0.137)	(0.182)
target: Voronezh	−0.026	−0.026	−0.026
	(0.052)	(0.052)	(0.052)
EI (corr. not shown)	−0.074	−0.046	−0.056
	(0.119)	(0.118)	(0.117)
EI (corr. shown)	−0.310*	−0.330*	−0.332*
	(0.134)	(0.132)	(0.132)
FIC	−0.301**	−0.283**	−0.272**
	(0.102)	(0.101)	(0.101)
target:voronezh × EI (corr. not shown)	0.005	0.006	0.005
	(0.087)	(0.087)	(0.087)
target:voronezh × EI (corr. shown)	0.377***	0.379***	0.378***
	(0.097)	(0.097)	(0.097)
target:voronezh × FIC	0.284***	0.285***	0.285***
	(0.074)	(0.074)	(0.074)
Age		0.046	0.026
		(0.033)	(0.035)
Gender		0.068	0.048
		(0.077)	(0.078)
CG first		0.010	0.000
		(0.077)	(0.077)
Head reported in CG		0.394***	0.398***
		(0.083)	(0.083)
Education			0.054
			(0.037)
Marital status			0.041
			(0.035)
Employment status			−0.024
			(0.020)
Income			0.009
			(0.035)
Knowledge index			0.032
			(0.047)
AIC	−1261.6	−1276.7	−1273.3
BIC	−1210.7	−1205.5	−1176.6
Log.Lik.	640.809	652.345	655.642

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Beta GLMM regression models. DV: honesty beliefs about specific region. Baseline region: Arkhangelsk. Baseline treatment: *FIN* (no corruption information available). CG - coin-flipping stage.

Treatment	Target region 1	Target region 2	n1	statistic	p.adj	signif
FIN	Arkhangelsk	Voronezh	200	20761.5	0.504	ns
FIC	Arkhangelsk	Voronezh	201	22933.5	0.017	*
EI (corr. shown)	Arkhangelsk	Voronezh	82	4165.0	0.008	**
EI (corr. not shown)	Arkhangelsk	Voronezh	117	6770.0	0.885	ns

Table 3: Paired Wilcoxon rank sum tests for the difference in first mover transfers in the trust game

Similar to the differences in belief in honesty in two regions with high and low levels of corruption, we calculated the variable *trust_diff*, defined as the difference in transfers made by first-movers in a trust game to a potential partner in Voronezh (presumably highly corrupt city) and Arkhangelsk (low corrupt city). The gaps between the Voronezh and Arkhangelsk transfers across treatments are shown in the right panel of Figure 1.

In their first mover transfer decisions to the *target* pool, participants from a *source* pool who observed regional corruption information (*FIC* treatment) sent 6.97 cents less to partners in the Voronezh region than to those in Arkhangelsk. Participants in no-corruption treatment sent 2.12 points less to Voronezh partners. This difference in gaps although is not statistically significant (Mann-Whitney-Wilcoxon test, MWW: $p = 0.105$).

This difference in gaps became larger and statistically significant if we look at the decisions made by those who voluntarily chose to observe corruption index in *EI* treatment and those who did not. Observers sent 11.46 less to Voronezh, while non-observers sent to Voronezh even 0.76 cents more than to Arkhangelsk (MWW: $p = 0.022$).

Examining trustor decisions toward participants from specific regions (Figure 3), we find that participants transferred an average of 46 cents out of 100 cents to participants from Voronezh (a more corrupt city), which is 13% less than to potential second movers from Arkhangelsk when corruption information was given in the *FIC* treatment (WRS: $p = 0.017$). In the FIN treatment without corruption, they sent 50 cents to Voronezh (just 4% less than to Arkhangelsk) (WRS: $p = 0.504$). Detailed results of the paired rank sum tests for each treatment are available in Table 3.

Similar to what we observed when we analysed the effect of corruption on beliefs, we found that endogeneity increases the effect of corruption information. In the *EI* treatment, participants who chose to observe corruption index transferred to those in Voronezh (a more corrupt city) 21% less than to potential second movers from

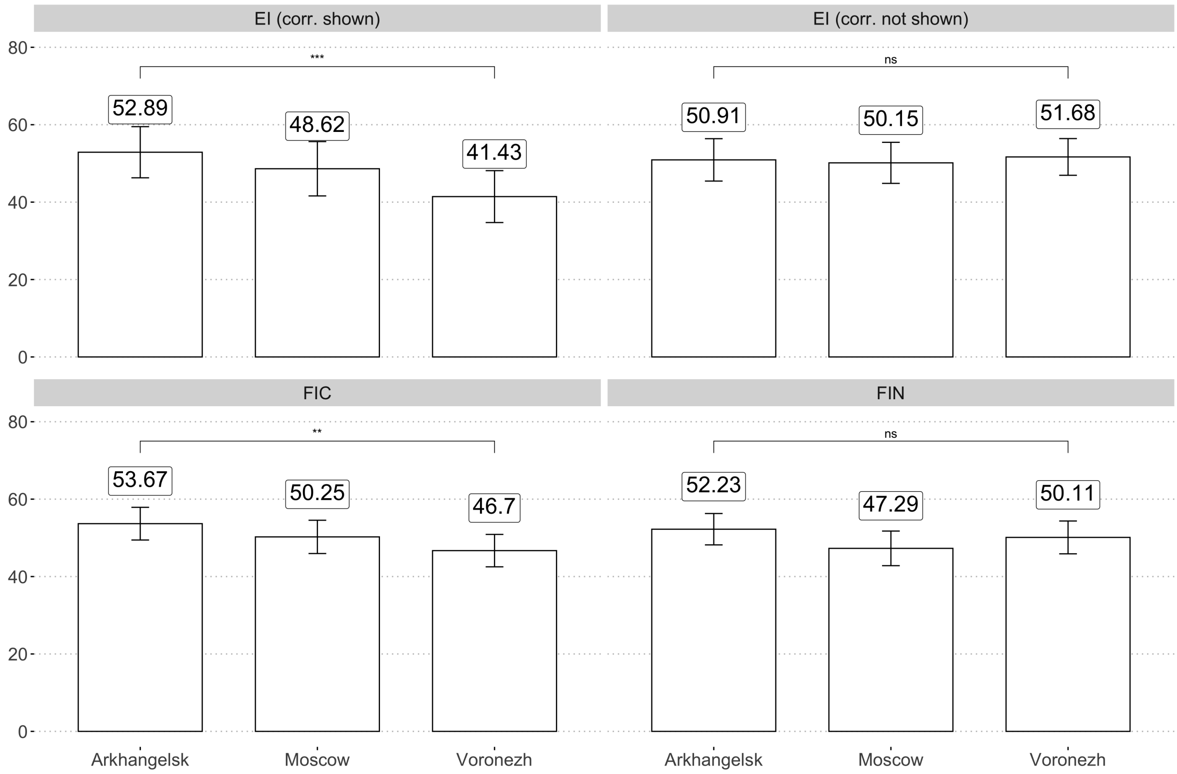


Figure 3: **Average transfers by first movers in a Trust game**. First movers (*Trustors*) decided how much out of 100 cents they send to their potential partners (*Trustees*) living in one of three cities (Arkhangelsk, Moscow, Voronezh). The whiskers show 95% confidence intervals, labels show the means. Mean comparisons between Arkhangelsk and Voronezh show results of paired samples Wilcoxon Rank Sum tests and symbols indicate statistical significance: ns: p > 0.05, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001. *FIC*– treatment with a fixed information set including regional corruption. *FIN*– treatment with a fixed information set without regional corruption. *EI*– endogenous information treatment where participants were able to choose information themselves.

Arkhangelsk (WRS: $p = 0.008$). They sent 1.5% more to Voronezh (WRS:$p = 0.885$) when they did not select corruption index.

The results of beta regressions are shown in Table 4. They confirm that trust decisions toward Voronezh participants are lower than toward Arkhangelsk participants in the *EI-corr* treatment. For the *FIC* treatment, this difference is barely significant.In extended models (Models 2 and 3) in which we control for game order, and reporting decisions in the *CG* stage, both these factors are significant for the trusting decisions.

H3 *Participants' estimates of the honesty of people from their own region are different when we give them the information about the corruption level of their own region*

H4 *The transfers of first movers in a trust game toward people from their own region are different when they are additionally provided with information about the corruption level of their own region.*

As the following analysis shows, neither the H3 nor the H4 hypothesis is supported by the available data. We have combined these two hypotheses because of their similarity from an analytical point of view. As Figure 4 shows, there are no differences between treatments in either honesty beliefs or trust decisions. Nonparametric tests of mean differences for CG (Kruskal-Wallis (KW) test, KW statistic = 5.72, p = 0.126) or first-mover transfers in a trust game (KW statistic = 1.17, p = 0.759) also showed no difference. The results of the pairwise Mann-Whitney and Wilcoxon tests are shown in the Appendix (Table A6).

4.1 Endogenous information

In the *EI* treatment, participants could choose three of six indicators to observe for each of the three regions examined in the study. Their preferences are shown in Figure 5: The corruption indicator ranked fourth in popularity and was chosen by 82 of 199 participants (41%). We also looked for two indicators that might shed light on why participants observe the corruption index per region: whether they first played the coin-flipping or trust game stages (this might tell us which of the two dimensions, honesty or trust corruption, might be more important), and whether those who self-reported heads were more likely to observe corruption. It should be noted that the population sizes in the subgroups within the EI treatments are relatively small (the total population is 200, and the number of individuals who chose to observe the corruption index is 82), so it is difficult to draw conclusions about the factors that influenced the decision to use the corruption index among the regional indicators based on the available data (see details in the Appendix, Figure A6).

	Model 1	Model 2	Model 3
(Intercept)	0.161	0.504*	0.668*
	(0.106)	(0.196)	(0.261)
target:Voronezh	−0.115	−0.115	−0.115
	(0.092)	(0.092)	(0.092)
subtreatment: EI (corr. not shown)	−0.121	−0.123	−0.122
	(0.173)	(0.172)	(0.172)
subtreatment: EI (corr. shown)	0.021	0.024	0.032
	(0.196)	(0.194)	(0.195)
subtreatment:FIC	0.074	0.050	0.047
	(0.149)	(0.149)	(0.149)
target:Voronezh × subtreatment:EI (corr. not shown)	0.183	0.181	0.180
	(0.153)	(0.153)	(0.153)
target:Voronezh × subtreatment:EI (corr. shown)	−0.434*	−0.432*	−0.430*
	(0.173)	(0.173)	(0.173)
target:Voronezh × subtreatment:FIC	−0.226+	−0.230+	−0.229+
	(0.132)	(0.132)	(0.132)
Age		−0.045	−0.047
		(0.047)	(0.050)
Gender		−0.276*	−0.261*
		(0.109)	(0.111)
CG first		0.200+	0.206+
		(0.109)	(0.109)
Head reported in CG		−0.267*	−0.274*
		(0.118)	(0.119)
Education			−0.023
			(0.053)
Marital status			0.005
			(0.050)
Employment status			0.000
			(0.028)
Income			−0.040
			(0.050)
Knowledge index			−0.042
			(0.067)
Num.Obs.	1200	1198	1198
R2 Marg.	0.018	0.051	0.054
R2 Cond.	1.008	1.008	1.008
AIC	−428.6	−436.0	−427.5
BIC	−377.7	−364.8	−330.8
ICC	1.0	1.0	1.0
RMSE	0.14	0.14	0.14

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Beta GLMM regression models. DV: transfers from a first mover to a partner in a given region. Baseline region: Arkhangelsk. Baseline treatment: *FIN* (no corruption information available). CG - coin-flipping stage.

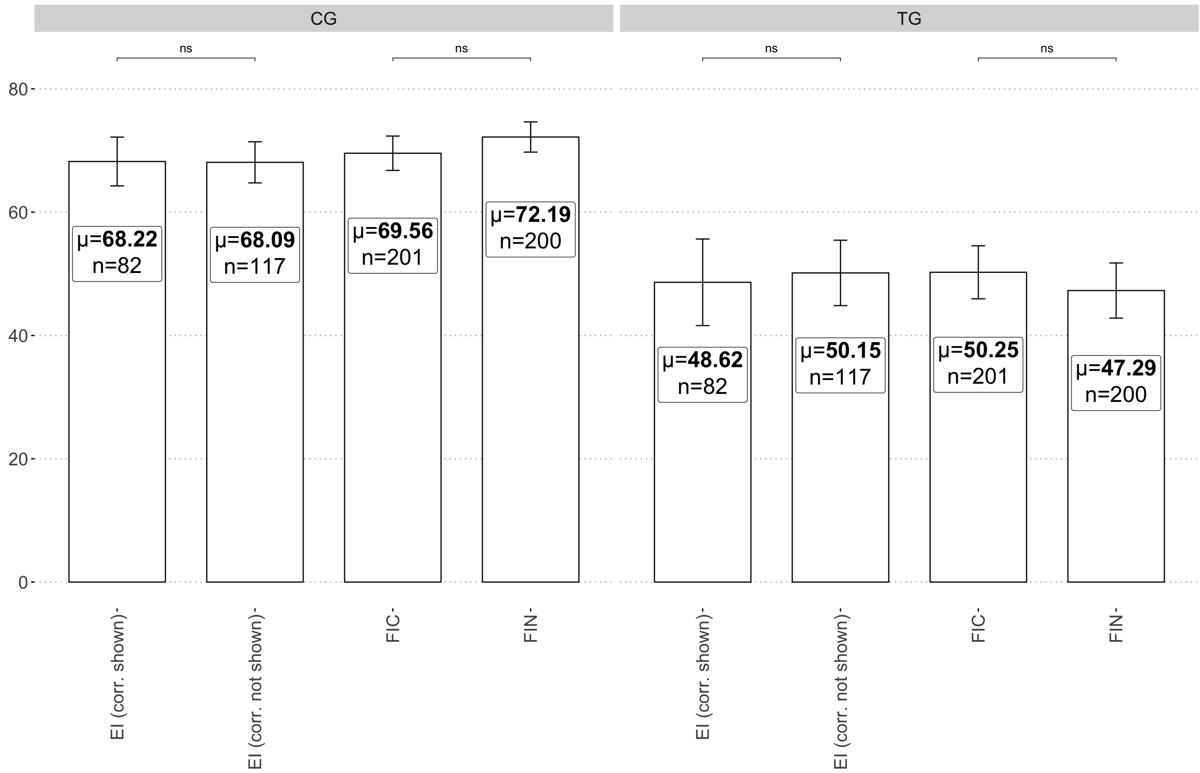


Figure 4: **Overall decisions about Moscow in *CG* and *TG***. *CG* denotes beliefs of a share (from 0 to 100%) of participants from Moscow who will report head in a cheating game. *TG* denotes transfers (from 0 to 100 cents) of first movers (*Trustors*) towards potential partners in Moscow. The whiskers show 95% confidence intervals, labels show the means (μ) and number of observations (n). Group comparisons show results of Mean comparisons show results of Mann-Whitney tests, *ns* indicates p > 0.05. *FIC*– treatment with a fixed information set including regional corruption. *FIN*– treatment with a fixed information set without regional corruption. *EI*– endogenous information treatment where participants were able to choose information themselves.

city	name	n	statistic	df	p	signif
Voronezh	CG	600	0.6771746	3	0.8790	
Voronezh	TG	600	9.3488872	3	0.0250	*
Arkhangelsk	CG	600	11.2362295	3	0.0105	*
Arkhangelsk	TG	600	0.6475541	3	0.8850	

Table 5: Kruskal-Wallis tests for Voronezh and Arkhangelsk. *CG* is a difference in beliefs about honesty of others in a coin-flipping game; *TG* is a difference in transfers by first movers in a trust game.

Finally, we analyzed the differences in honesty beliefs and trust decisions in Arkhangelsk and Voronezh separately. This may shed light on whether or not the observation of the corruption index causes people to reconsider their attitudes toward more or less corrupt regions. As the results of the Kruskal-Wallis tests show (Table 5), for Arkhangelsk there is a difference in honesty beliefs between treatments, but no difference in initial decisions in a trust game. In contrast, trust decisions in Voronezh differ statistically between treatments but are indistinguishably the same in terms of beliefs. The additional pairwise tests and graphs for individual cities are available in the Appendix (figures A12, A11; tables A7, A8). These results may suggest that people who are exposed to information about corruption begin to trust residents of more corrupt cities less, but do not trust residents of less corrupt cities more. The reverse is true for estimates of honesty: people reconsider their estimates of the honesty of residents of less corrupt regions, but this does not happen for estimates of honesty for less corrupt regions.

5 Discussion

In this paper we find evidence that information about corruption in a region can affect interpersonal relations in quite unforeseen ways. As we have seen, this information can undermine trust and make people believe that others are less honest than they would have thought. The most worrisome interpretation of this finding is that people are easily susceptible to the manipulation of information that pushes them towards statistical discrimination. In particular, people are quick to draw conclusions about the honesty and trustworthiness of a single person or a small subgroup from an indicator that applies to an entire region.

Although information on regional corruption may trigger statistical discrimination, the question remains whether it is really that harmful. What if there is a reason for it? In other words, are people from more corrupt regions actually less trustworthy and more dishonest? The evidence is mixed. Country of origin corruption level was a good predictor

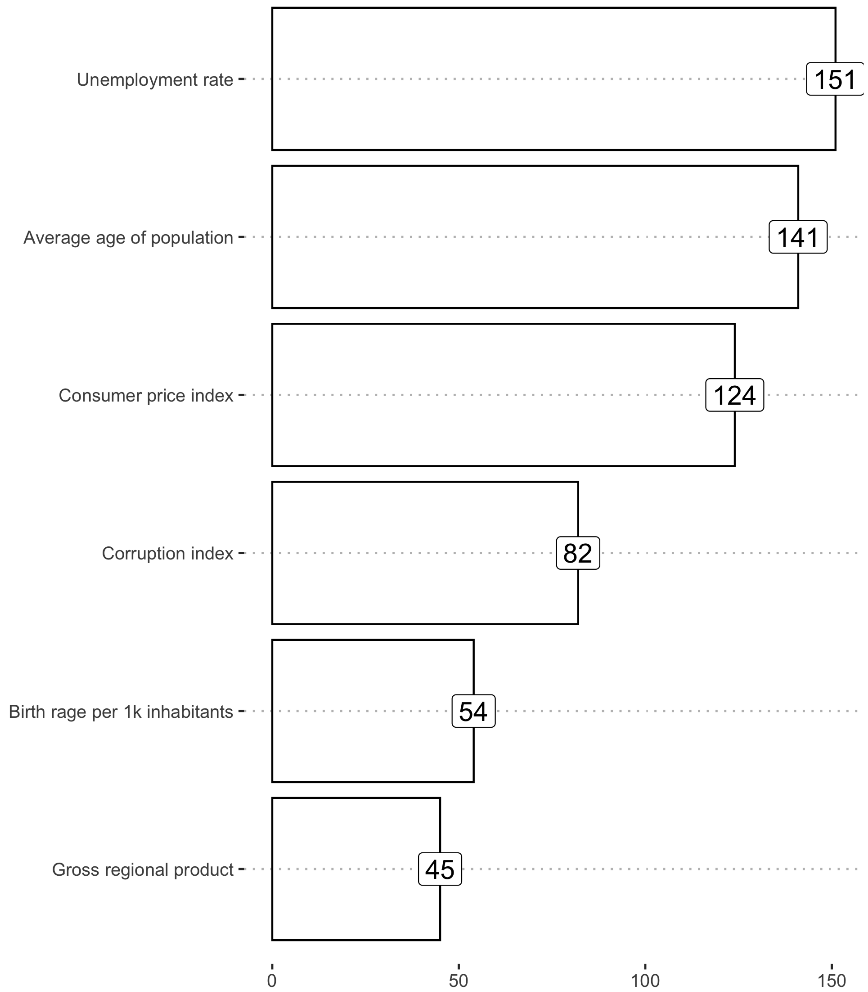


Figure 5: How often the indicators were chosen in *EI* treatment. Only participants in *EI* (endogenous information) treatment ($N = 199$) are counted. Each participant had to choose 3 out of 6 indicators. Order of indicators was randomized. Labels show number of participants who chose an indicator.

of behavior in the corruption game among students in the United Kingdom, so there is some basis for those who fall for statistical discrimination. However, this predictor failed for graduate students and gradually disappeared the longer a person lived in the United Kingdom [Barr and Serra, 2010]. Taking advantage of the fact that southern Italian regions are perceived as more corrupt than the rest of the country, the opposite effect was shown: participants from the south of Italy were both less willing to offer bribes and to believe that their counterpart was corrupt [Zhang, 2015].

This paper is not the first to examine the effects of information in general, and corruption in particular, on the decisions of participants in behavioral games. Information about the trustee's previous decisions [Bracht and Feltovich, 2009], partner's physical attractiveness [Wilson and Eckel, 2006], and even whether the partner's name was easy or difficult to pronounce [Zürn and Topolinski, 2017] all contributed to the degree of trust that a trustor exhibited. To our knowledge, however, the effect of perceived corruption of the partner's region on the partner's trustworthiness and honesty has not been studied before.

The design of the study and the online population from which the data were collected limit the generalizability and external validity of the study. First, the percentage of actual cheating and beliefs about cheating should be viewed with caution because an online audience tends to cheat more frequently than one in a laboratory and is generally less trusting [Dickinson and McEvoy, 2021]. Such context-depending subjects as honesty and trust are difficult to measure adequately in the laboratory or in online environments. It would be particularly important to see if the deleterious effect of corruption information on trust carries over to real life. In particular, whether it actually affects the life chances of those who live in or come from presumably more corrupt regions? Are people from there less likely to get a job because they are seen as less credible? Are traffic offenders crossing the country more likely to offer a bribe to a road policeman in a region they consider more corrupt? Then corruption would indeed become a self-fulfilling prophecy, as Corbacho et al. [2016] aptly put it.

Another important limitation of the study is the weak prejudices that people usually have towards other regions within their own country. Unlike national or ethnic bigotry, prejudice against a particular region (when it does not coincide with the settlement of a particular ethnic or racial group) can be quite weak. The stronger the pre-existing expectations regarding certain characteristics of the area's inhabitants, the less likely it is that a simple nudge can significantly change them. This could explain why we did not observe shifts in honesty and trust in the home region between treatments. It could also mean that the impact of country-level corruption indicators is less harmful than can be inferred from our work, because the perception of individual countries should be stronger than that of regions.

The most promising area for further investigation is what influences the demand for endogenous selection of information. When deciding to choose an available regional information to properly assess the overall honesty and trustworthiness of participants from each region, less than half of the participants chose the corruption index. One of the key factors could be the demand for self-serving information: Those who chose to cheat preferred to find some indicators of the dishonesty of others. However, it remains an open question what kind of group-level information they would actually consider crucial for assessing individual dishonesty and trust. Substantially expanding the list of available regional statistics from which they can choose, and estimating their willingness to pay for such information, might provide us with insights into a more realistic description of statistical discrimination of this type in the future. Thus, this study is an important, but only a first step toward a more sophisticated examination of the unintended consequences of corruption indices or regional statistics in general.

A Appendix

A.1 Balance tests

Table A1: Summary Statistics for the *Source* pool

treatment	ei			fic			fin			
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	Test
age	199	2.678	1.213	200	2.79	1.18	200	2.555	1.064	F=2.074
education	199	2.437	1.071	200	2.24	1.108	200	2.285	1.034	F=1.854
gender	199			200			200			X2=4.399
... Female	104	52.3%		84	42%		98	49%		
... Male	95	47.7%		116	58%		102	51%		
marital	199	1.241	1.252	200	1.05	1.036	200	1.065	1.148	F=1.709
employment	199	2.141	1.995	200	1.92	2.053	200	1.935	1.929	F=0.763
income	199	2.799	1.064	200	2.795	1.118	200	2.71	1.18	F=0.401
instructions_clarity	199	4.357	0.869	201	4.313	0.828	200	4.41	0.771	F=0.691
general_risk	199	4.643	2.172	200	4.65	2.246	200	4.67	2.286	F=0.008
general_trust	199	1.774	0.419	200	1.845	0.363	200	1.81	0.393	F=1.638
religion	199	4.583	3.002	200	5.01	2.858	200	4.665	2.985	F=1.18
political	199	5.075	1.738	200	5.455	1.851	200	5.105	1.743	F=2.817
arkh_knowledge_index	199	1.221	0.927	201	1.224	0.951	200	1.235	0.946	F=0.012
voronezh_knowledge_index	199	1.407	0.985	201	1.318	1.048	200	1.465	1.007	F=1.064
moscow_knowledge_index	199	1.603	1.163	201	1.831	1.331	200	1.6	1.169	F=2.348
knowledge_diff	199	0.186	0.667	201	0.095	0.846	200	0.23	0.923	F=1.427

Statistical significance markers: * p<0.05; ** p<0.01; *** p<0.001

A.2 Screenshots

A.3 Flowchart of screens

Source: [FOM, 2011]

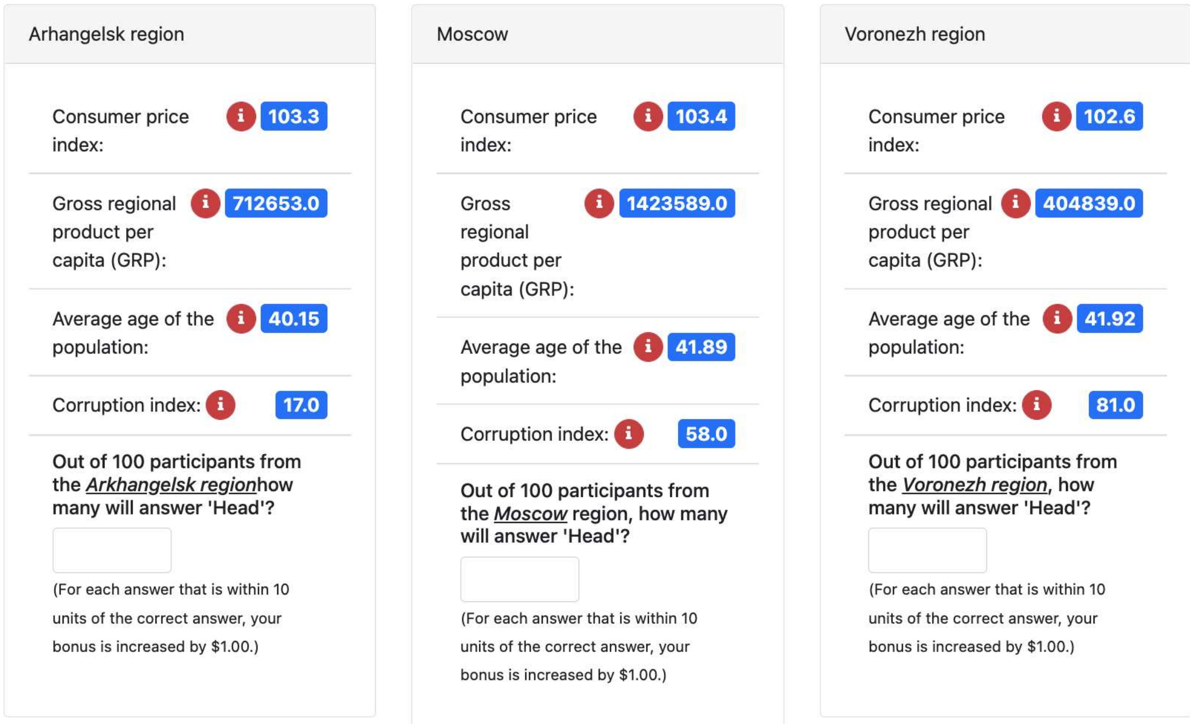


Figure A1: Screenshot of the decision stage of CG game, *FIC* treatment (automatic translation from Russian)

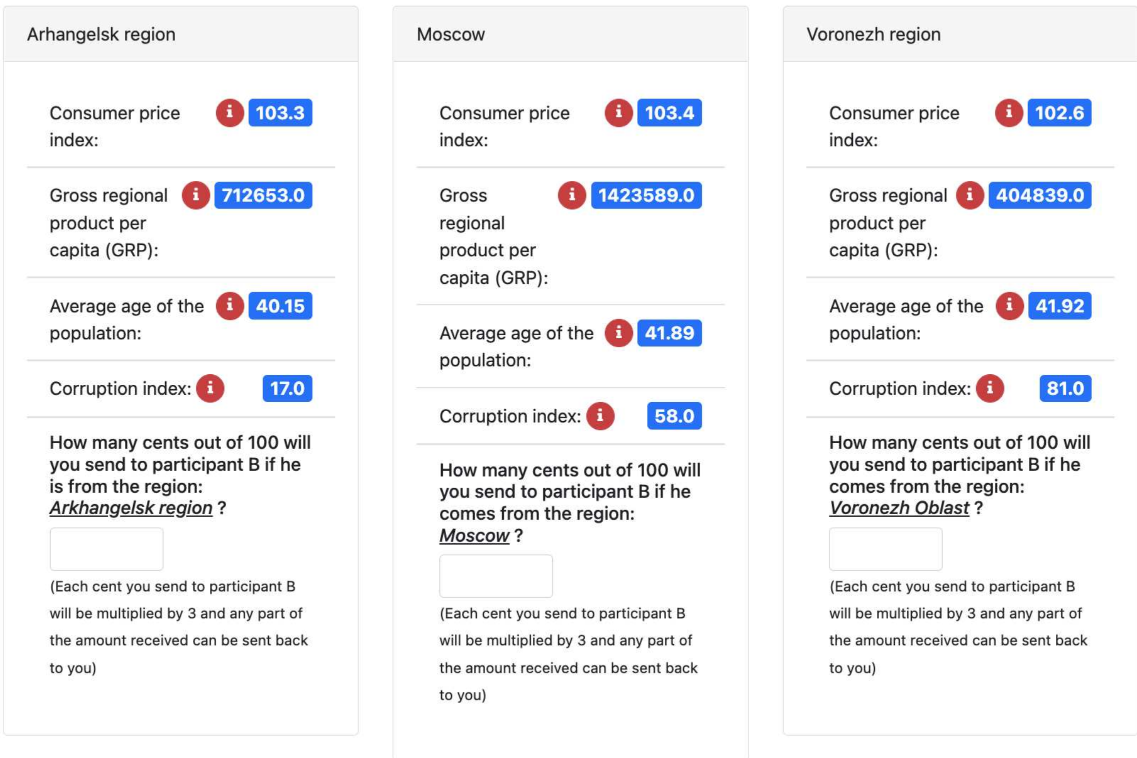
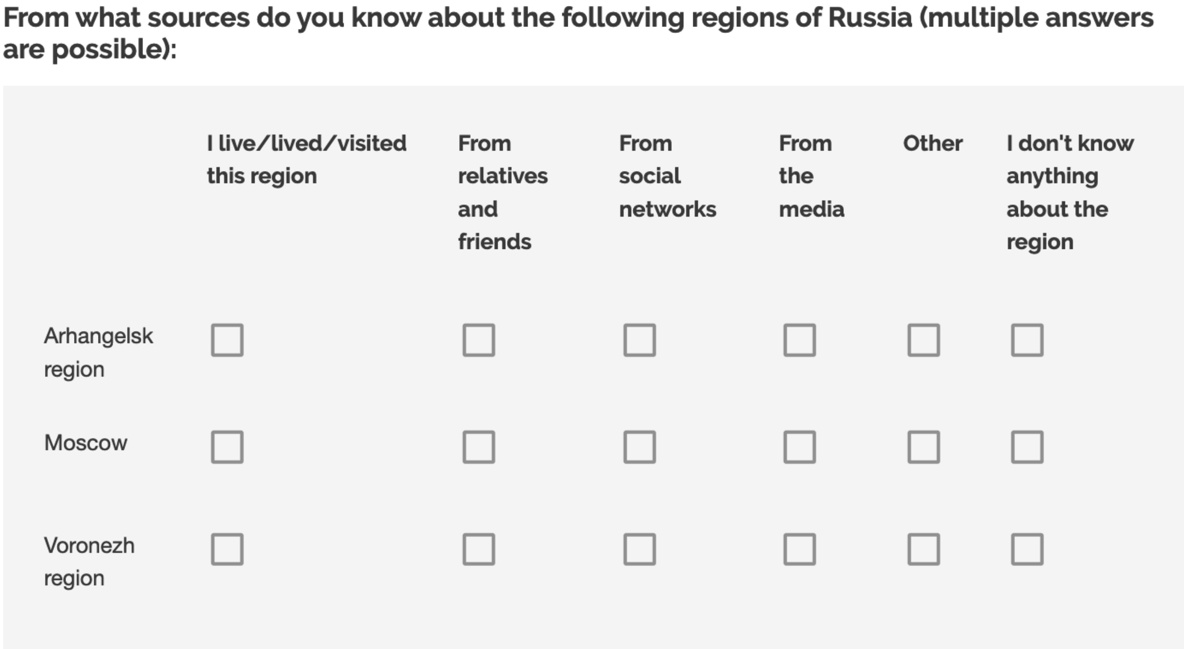


Figure A2: Screenshot of the decision stage of TG game, *FIC* treatment (automatic translation from Russian)



From what sources do you know about the following regions of Russia (multiple answers are possible):

	I live/lived/visited this region	From relatives and friends	From social networks	From the media	Other	I don't know anything about the region
Arhangelsk region	☐	☐	☐	☐	☐	☐
Moscow	☐	☐	☐	☐	☐	☐
Voronezh region	☐	☐	☐	☐	☐	☐

Figure A3: Questions regarding regional knowledge (automatic translation from Russian)

A.4 Effect of order and own decision in CG on preferences for information

A.5 Testing for normality

variable	statistic	p
arkh_cg_belief	0.9509197	3.14e-13
arkh_tg_decision	0.9340205	1.00e-15
belief_diff	0.8645156	0.00e+00
msk_cg_belief	0.9466821	7.10e-14
msk_tg_decision	0.9227373	0.00e+00
trust_diff	0.8487366	0.00e+00
voronezh_cg_belief	0.9385296	5.00e-15
voronezh_tg_decision	0.9332793	1.00e-15

Table A2: Results of Shapiro-Wilk Normality Tests

A.6 OLS models for differences in trust and beliefs

We do estimate the following model:

$$\Delta b = \beta_0 + \beta_1 T + \beta_2 O + \beta_3 C + \beta_4 \mathbf{X} + \epsilon_i \tag{4}$$

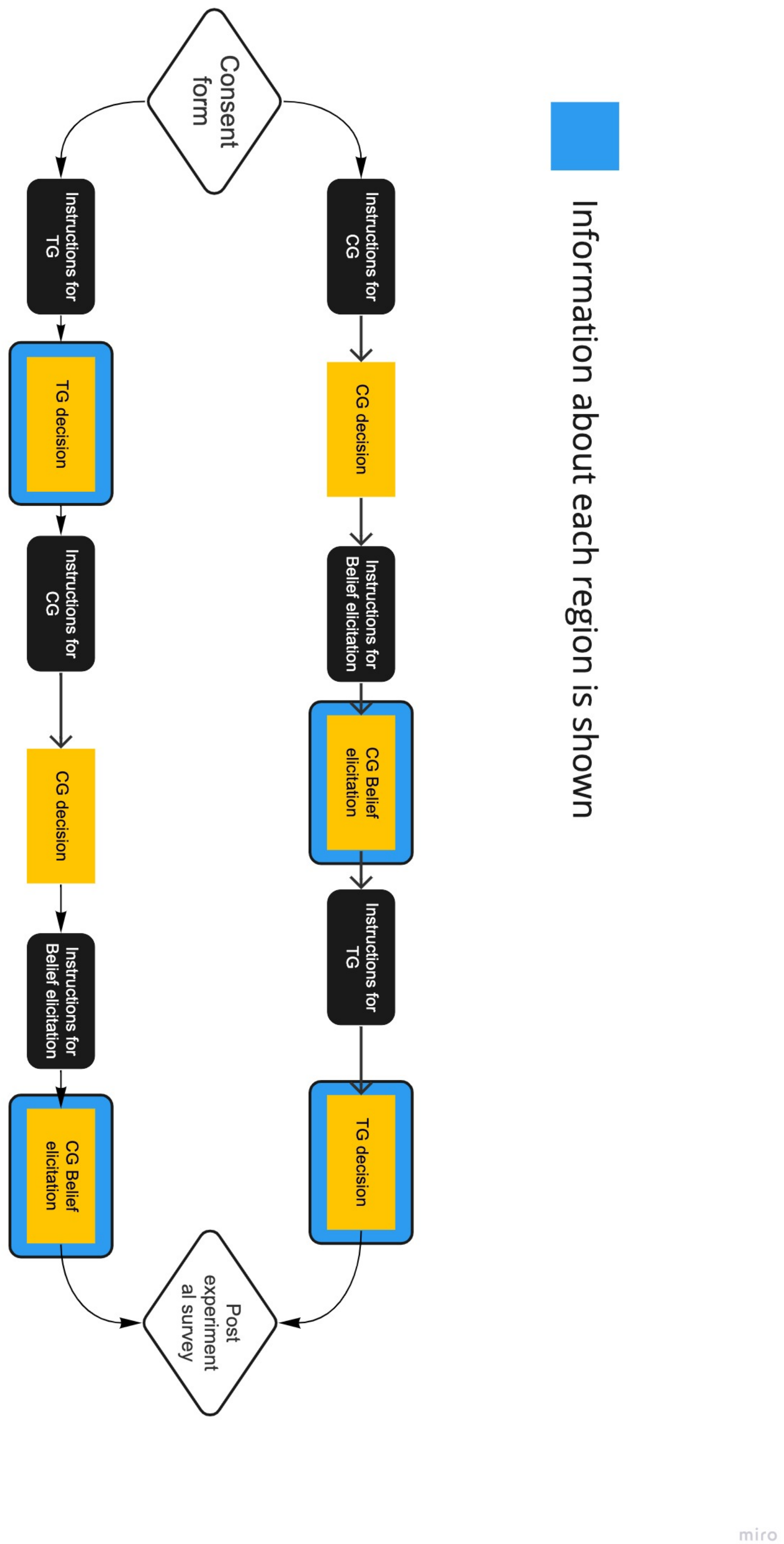
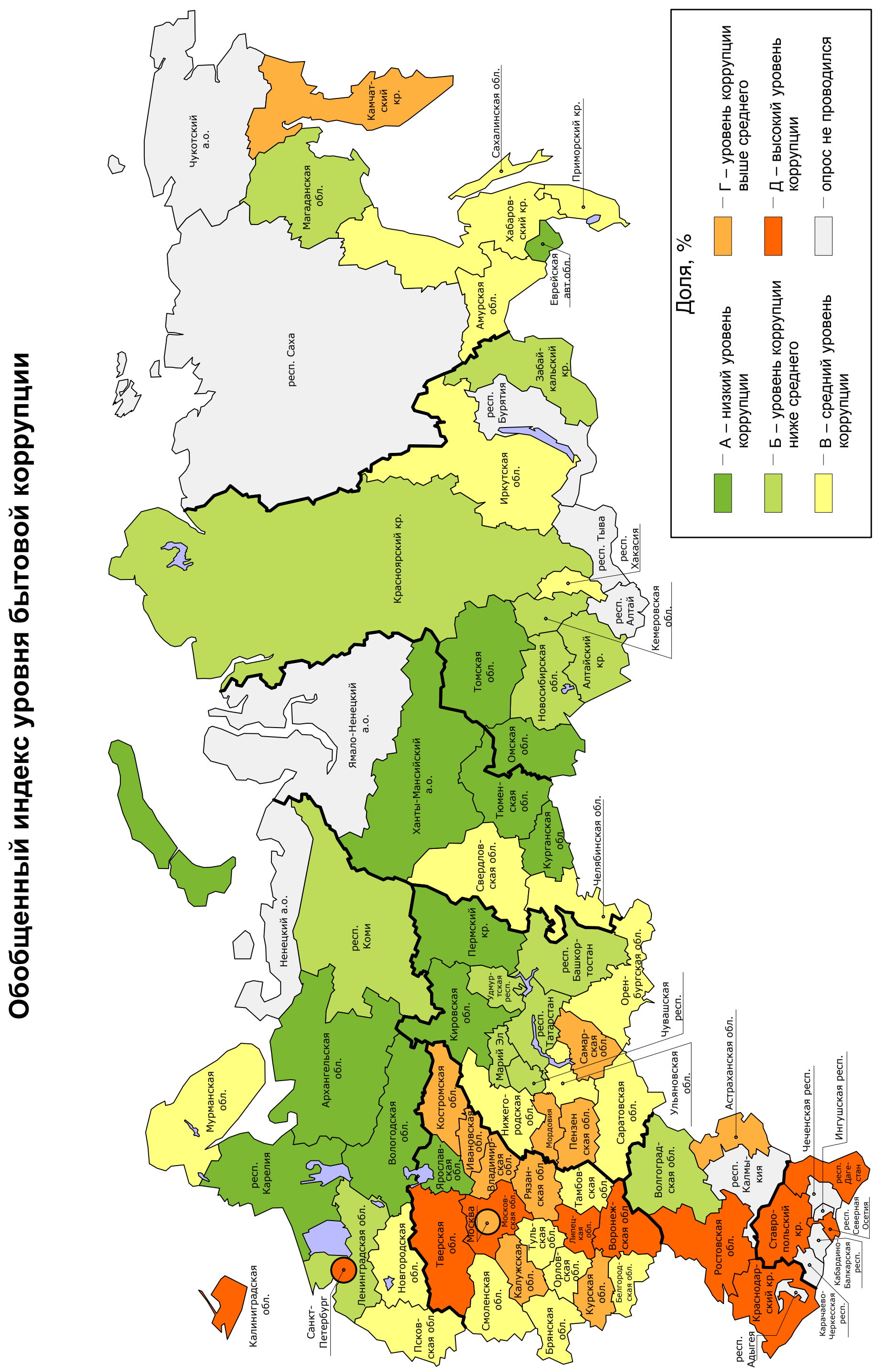


Figure A4: Flowchart of screens

37



Карта 2. Обобщенный индекс уровня бытовой коррупции

Figure A5: Map of regional corruption index. Source: FOM [2011]

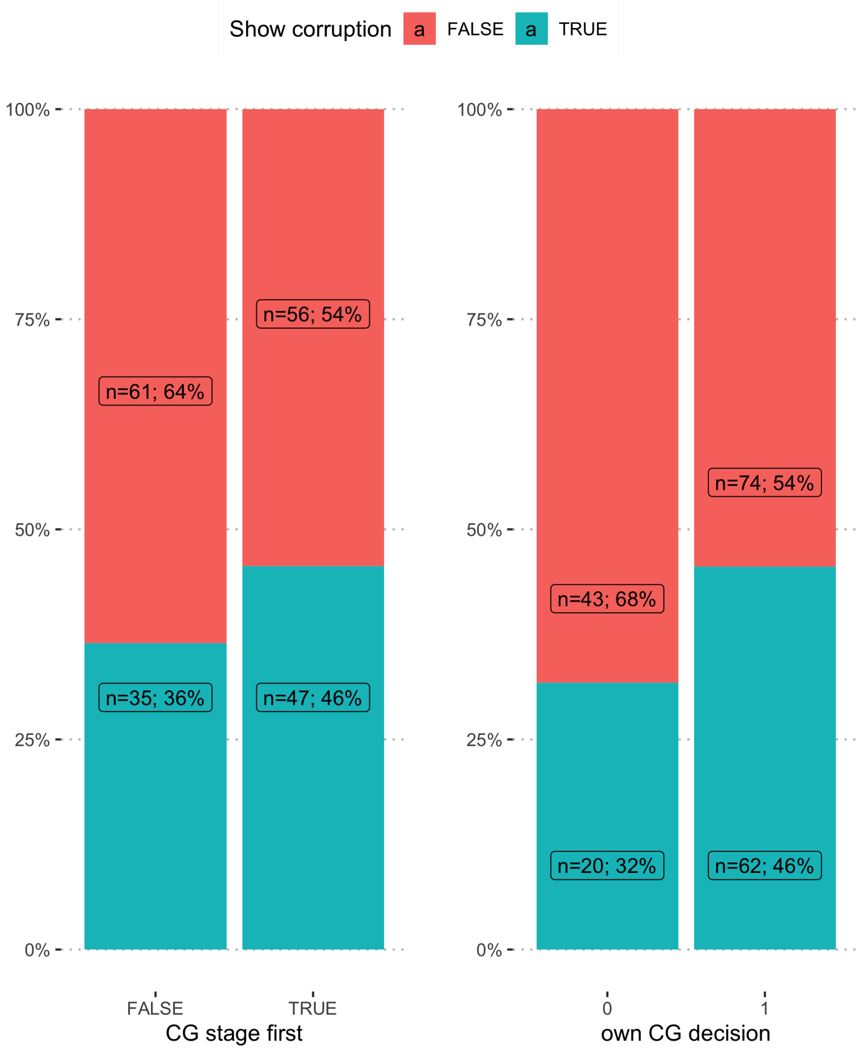


Figure A6: Decision to observe corruption by CG decision and game order

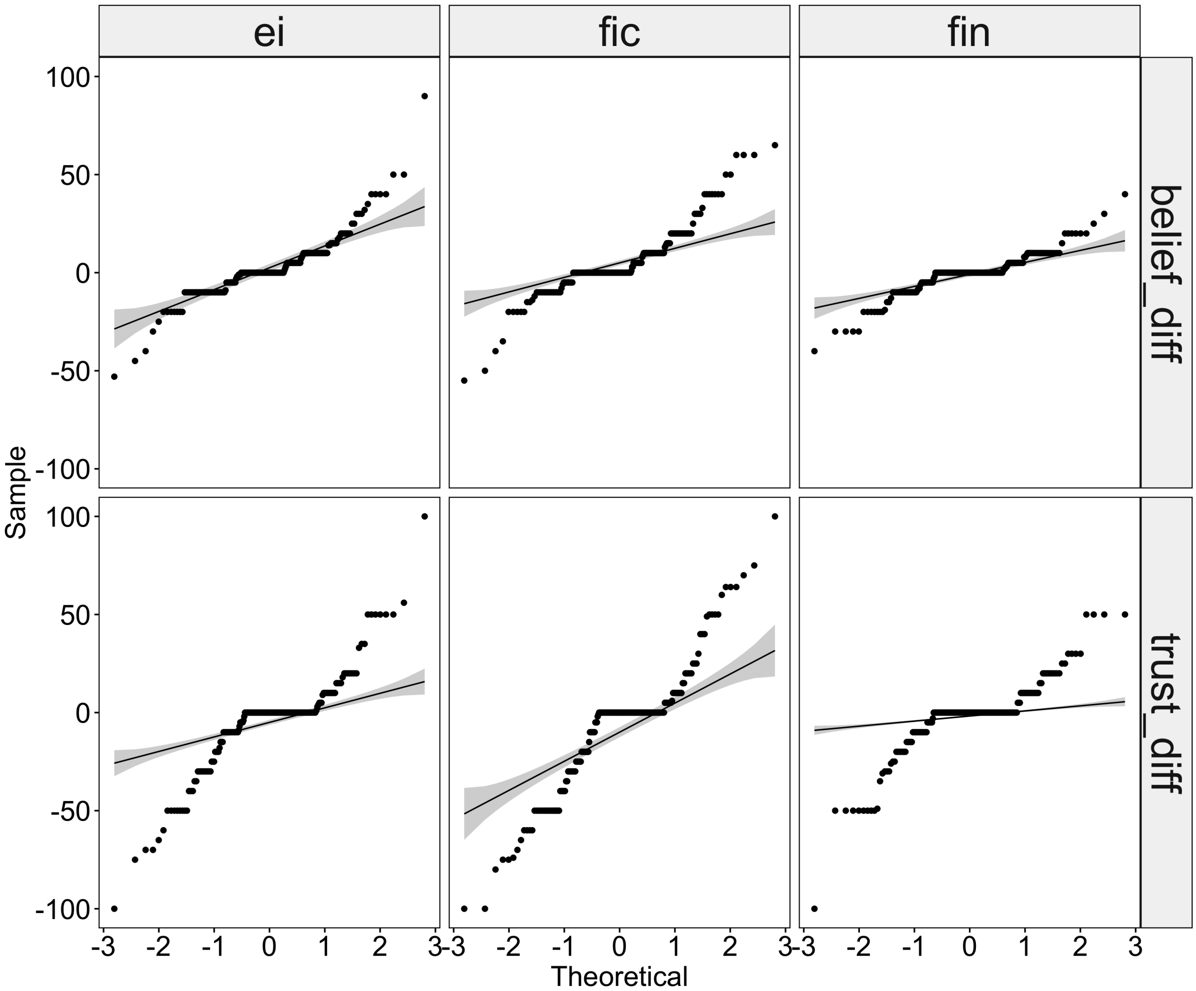


Figure A7: A quantile-quantile (Q-Q) plot of gaps in beliefs and trust. *belief-diff* is a difference at the individual level between beliefs about share of reporting heads in Voronezh and Arkhangelsk. *trust-diff* is a difference between transfers to partners in Voronezh and Arkhangelsk. *FIC*− treatment with a fixed information set including regional corruption. *FIN*− treatment with a fixed information set without regional corruption. *EI*− endogenous information treatment where participants were able to choose information themselves.

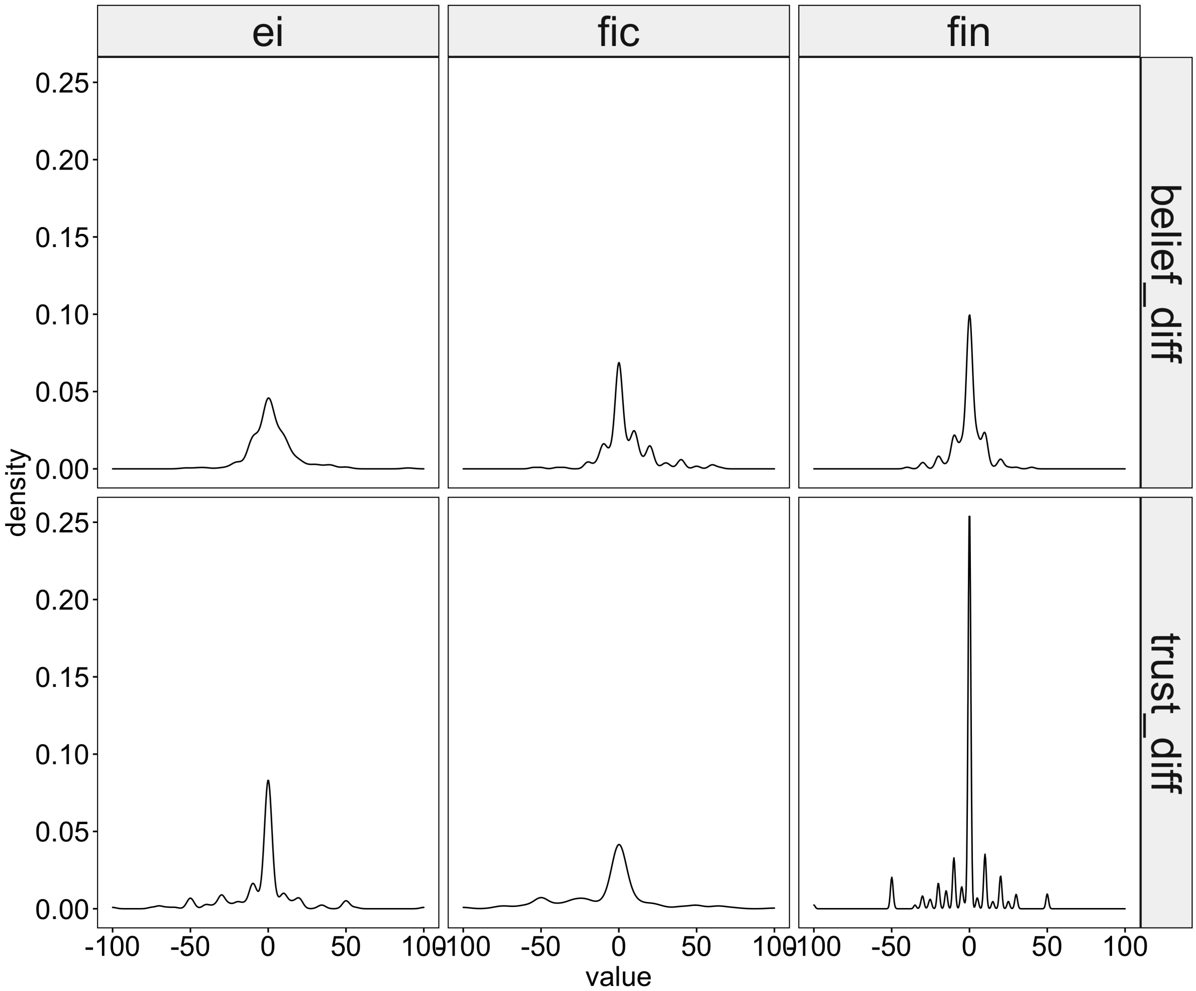


Figure A8: A density plot of gaps in beliefs and trust. *belief-diff* is a difference at the individual level between beliefs about share of reporting heads in Voronezh and Arkhangelsk. *trust-diff* is a difference between transfers to partners in Voronezh and Arkhangelsk. *FIC*– treatment with a fixed information set including regional corruption. *FIN*– treatment with a fixed information set without regional corruption. *EI*– endogenous information treatment where participants were able to choose information themselves.

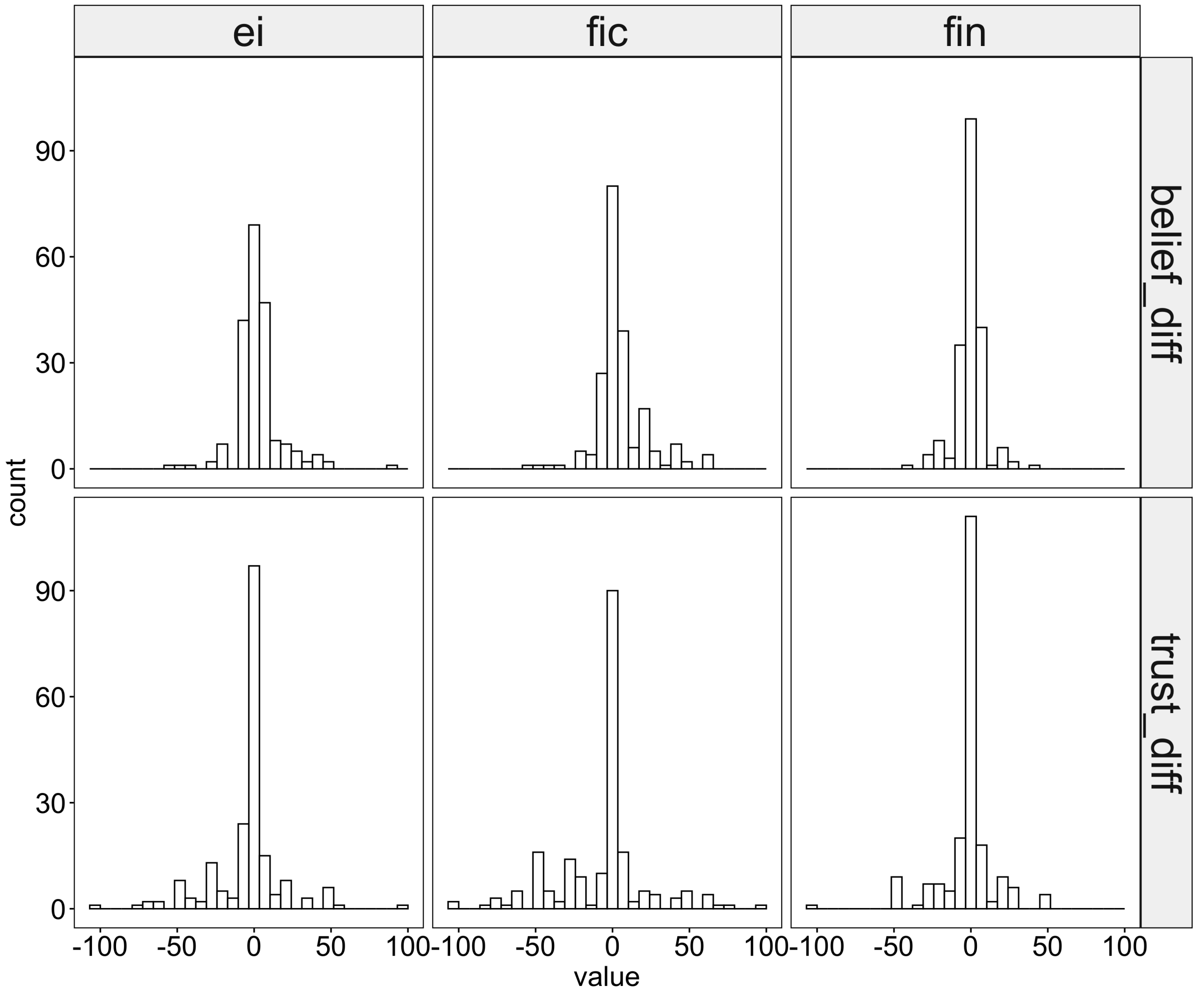


Figure A9: Histogram of gaps in beliefs and trust. *belief-diff* is a difference at the individual level between beliefs about share of reporting heads in Voronezh and Arkhangelsk. *trust-diff* is a difference between transfers to partners in Voronezh and Arkhangelsk. *FIC*− treatment with a fixed information set including regional corruption. *FIN*− treatment with a fixed information set without regional corruption. *EI*− endogenous information treatment where participants were able to choose information themselves.

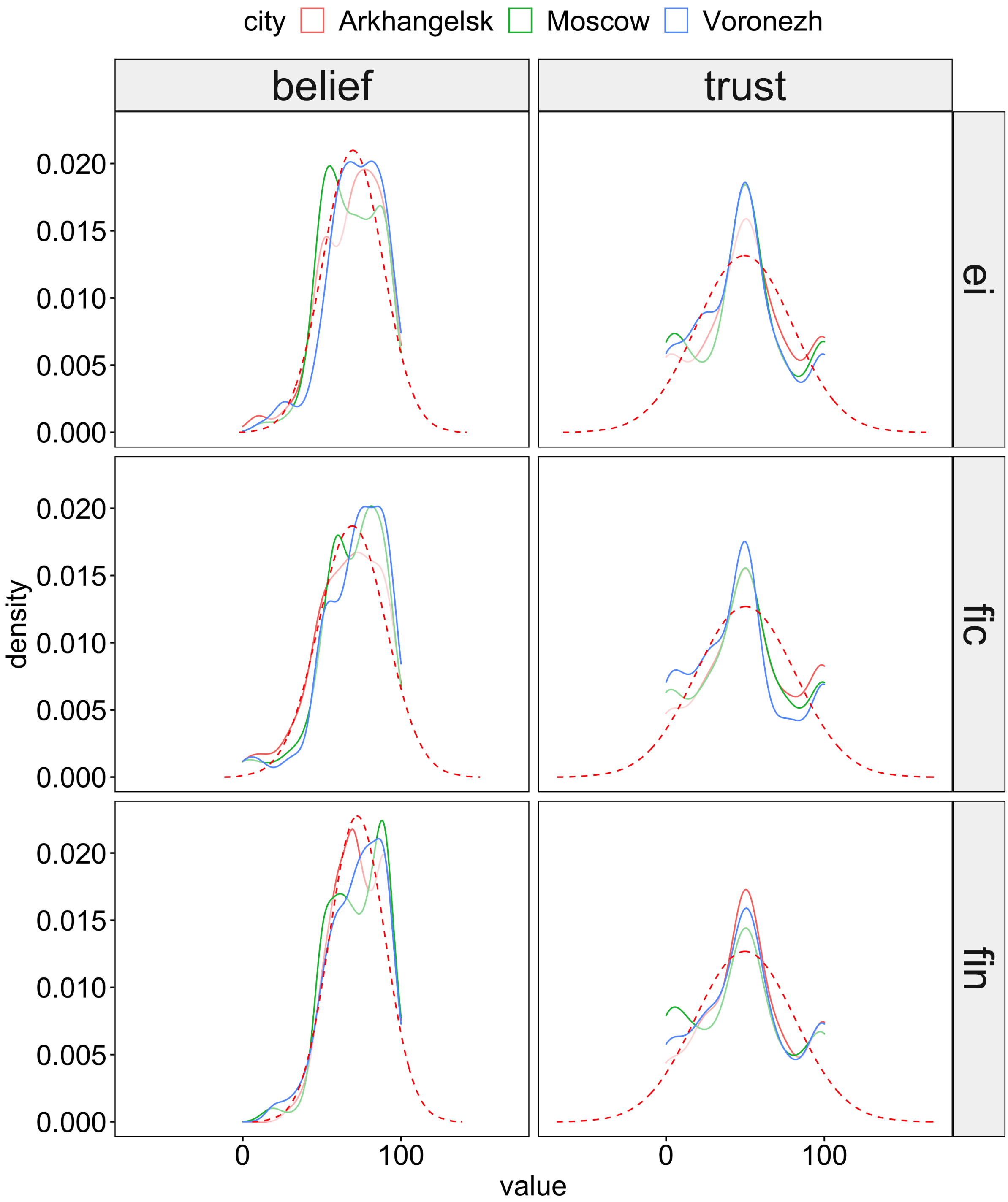


Figure A10: Density curves of individual beliefs about honesty and transfers by first movers in a Trust game towards single regions. *Belief* is an individual estimate of a share (out of 100 participants) of those who report a head after flipping a coin in each region. *Trust* is a first mover decision in a Trust game towards a potential partner in a specific region. *FIC*– treatment with a fixed information set including regional corruption. *FIN*– treatment with a fixed information set without regional corruption. *EI*– endogenous information treatment where participants were able to choose information themselves.

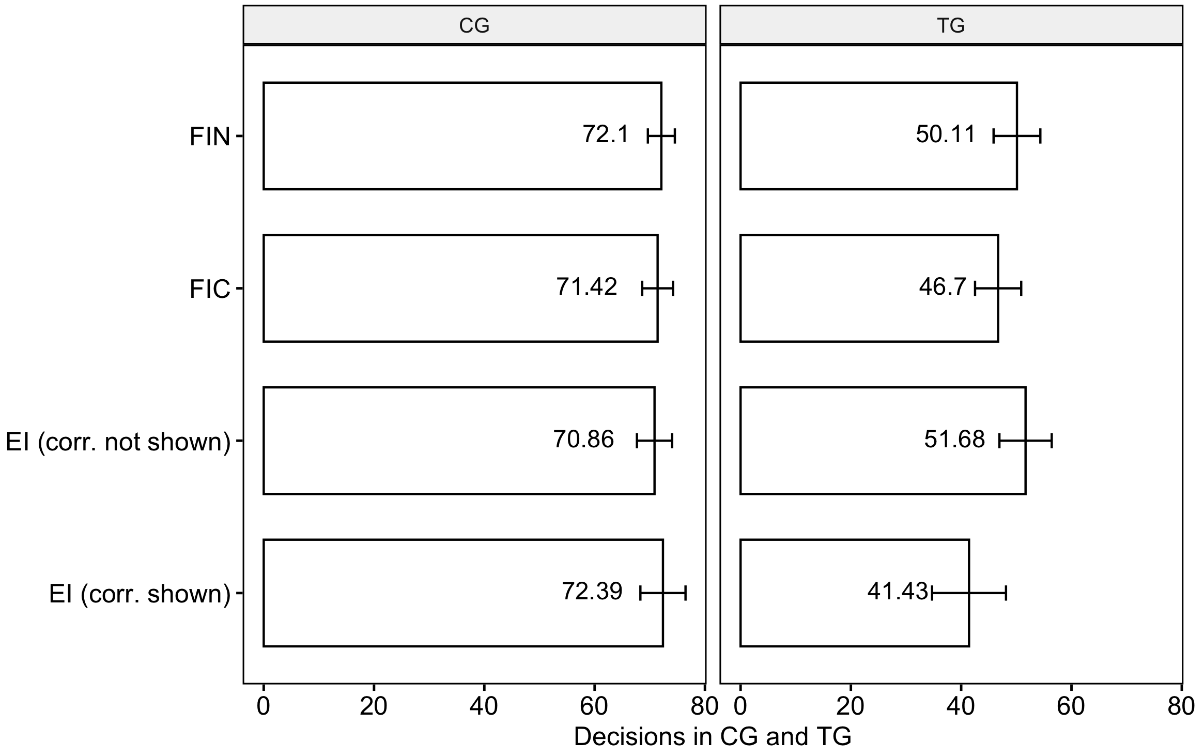


Figure A11: Overall decisions about Voronezh in *CG* and *TG*

Here, Δb is *belief_diff* variable, the difference between individual estimates of honesty (proportion who indicate heads in the coin flipping stage) in Voronezh and Arkhangelsk; T is a treatment effect; O is the order in which two stages (coin flippint or trust game) are played; C is an individual indicating his or her own coin flipping decision; and $\mathbf{X}$ is the vector of sociodemographic controls.

A.7 Results of paired Wilcoxon tests on beliefs across regions and treatments

	base	extended base	full	full+WVS
(Intercept)	−0.465	−0.315	0.584	0.807
	(1.000)	(2.398)	(2.702)	(4.852)
subtreatmentFIC	5.485***	5.461***	5.418***	5.370***
	(1.413)	(1.426)	(1.441)	(1.472)
subtreatmentEI (corr. shown)	7.038***	7.058***	6.925***	7.009***
	(1.855)	(1.862)	(1.881)	(1.908)
subtreatmentEI (corr. not shown)	−0.022	−0.230	−0.349	−0.226
	(1.646)	(1.656)	(1.686)	(1.716)
cg_firstTRUE		−0.163	−0.388	−0.402
		(1.162)	(1.178)	(1.200)
cg_decisionHead		−1.564	−1.626	−1.573
		(1.265)	(1.280)	(1.299)
knowledge_diff			0.579	0.454
			(0.720)	(0.737)
religion				−0.049
				(0.216)
political				−0.176
				(0.363)
Num.Obs.	600	599	599	599
R2	0.043	0.047	0.062	0.072
R2 Adj.	0.039	0.036	0.032	0.021
AIC	4888.0	4886.4	4900.9	4918.9
BIC	4910.0	4925.9	4993.2	5063.9
Log.Lik.	−2438.995	−2434.194	−2429.459	−2426.449
F	9.032	4.199	2.025	1.413
Age and gender:	NO	YES	YES	YES
Other soc.dem:	NO	NO	YES	YES
WVS, risk and trust:	NO	NO	NO	YES

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table A3: OLS regression results. **DV**: *belief_diff*, difference between beliefs about share of heads reported in Voronezh (high corrupt region) and Arkhangelsk (low corrupt region). Baseline treatment is *FIN*. Baseline *target* region is Arkhangelsk

	base	extended base	full	full+WVS
(Intercept)	−2.120	0.051	−1.748	5.464
	(1.711)	(4.498)	(4.996)	(8.690)
subtreatmentFIC	−4.845*	−4.517+	−4.670+	−5.066*
	(2.417)	(2.455)	(2.485)	(2.522)
subtreatmentEI (corr. shown)	−9.343**	−9.260**	−9.987**	−11.344***
	(3.173)	(3.215)	(3.258)	(3.283)
subtreatmentEI (corr. not shown)	2.881	3.065	2.599	2.414
	(2.816)	(2.858)	(2.922)	(2.948)
cg_firstTRUE		1.873	2.034	2.489
		(1.994)	(2.027)	(2.044)
cg_decisionHead		2.607	2.335	2.773
		(2.172)	(2.203)	(2.216)
knowledge_diff			−0.254	−0.010
			(1.242)	(1.261)
religion				−0.111
				(0.369)
political				0.087
				(0.622)
Num.Obs.	600	599	599	599
R2	0.027	0.039	0.049	0.074
R2 Adj.	0.022	0.019	0.009	0.014
AIC	5532.2	5534.6	5552.0	5560.4
BIC	5554.2	5596.1	5666.3	5727.4
Log.Lik.	−2761.123	−2753.305	−2750.021	−2742.201
F	5.446			
RMSE	24.12	23.99	23.86	23.55
Age and gender:	NO	YES	YES	YES
Other soc.dem:	NO	NO	YES	YES
WVS, risk and trust:	NO	NO	NO	YES

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4: OLS regression results. **DV**: *trust_diff*, difference between transfers by first movers in the Trust game towards potential partners in Voronezh (high corrupt region) and Arkhangelsk (low corrupt region). Baseline treatment is *FIN*. Baseline *target* region is Arkhangelsk

Table A5: paired Wilcoxon rank sum tests onthe difference in beliefs

subtreatment	group1	group2	n1	n2	statistic	p.adj	p.adj.signif
EI (corr. shown)	Arkhangelsk	Voronezh	82	82	2659.0	0.020	*
EI (corr. not shown)	Arkhangelsk	Voronezh	117	117	7110.0	0.606	ns
FIC	Arkhangelsk	Voronezh	201	201	17347.0	0.014	*
FIN	Arkhangelsk	Voronezh	200	200	19885.5	0.921	ns

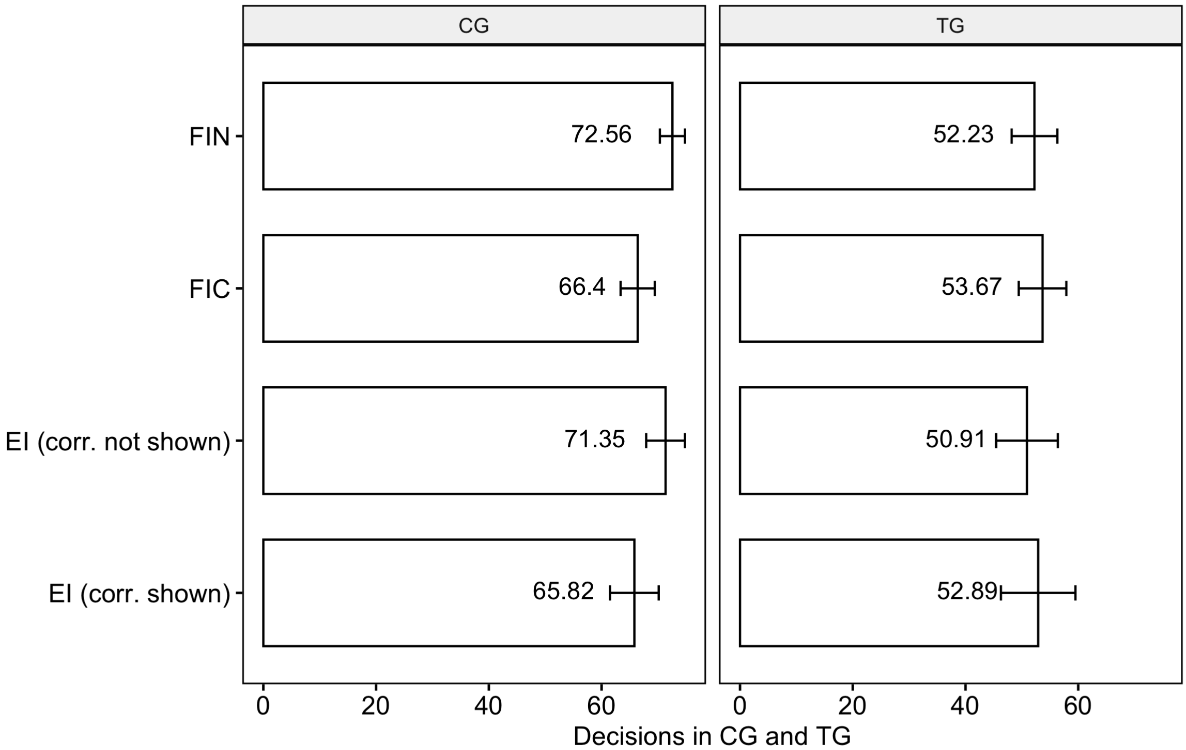


Figure A12: Overall decisions about Arkhangesk in *CG* and *TG*

Table A6: Pairwise Wilcoxon test for Moscow decisions

name	.y.	group1	group2	n1	n2	statistic	p	p.adj	p.adj.signif
CG	value	EI (corr. shown)	EI (corr. not shown)	82	117	4812.5	0.970	0.970	ns
CG	value	EI (corr. shown)	FIC	82	201	7562.5	0.275	0.944	ns
CG	value	EI (corr. shown)	FIN	82	200	7058.5	0.065	0.323	ns
CG	value	EI (corr. not shown)	FIC	117	201	10827.5	0.236	0.944	ns
CG	value	EI (corr. not shown)	FIN	117	200	10138.0	0.046	0.273	ns
CG	value	FIC	FIN	201	200	18928.0	0.309	0.944	ns
TG	value	EI (corr. shown)	EI (corr. not shown)	82	117	4636.5	0.683	1.000	ns
TG	value	EI (corr. shown)	FIC	82	201	8028.0	0.730	1.000	ns
TG	value	EI (corr. shown)	FIN	82	200	8418.5	0.722	1.000	ns
TG	value	EI (corr. not shown)	FIC	117	201	11832.5	0.925	1.000	ns
TG	value	EI (corr. not shown)	FIN	117	200	12386.5	0.376	1.000	ns
TG	value	FIC	FIN	201	200	21165.5	0.352	1.000	ns

Table A7: Pairwise Wilcoxon test for Arkhangelsk decisions

name	.y.	group1	group2	n1	n2	statistic	p	p.adj	p.adj.signif
CG	value	EI (corr. shown)	EI (corr. not shown)	82	117	3905.0	0.025	0.100	ns
CG	value	EI (corr. shown)	FIC	82	201	7886.5	0.569	1.000	ns
CG	value	EI (corr. shown)	FIN	82	200	6561.0	0.008	0.048	*
CG	value	EI (corr. not shown)	FIC	117	201	13317.0	0.048	0.143	ns
CG	value	EI (corr. not shown)	FIN	117	200	11700.0	1.000	1.000	ns
CG	value	FIC	FIN	201	200	17269.5	0.014	0.070	ns
TG	value	EI (corr. shown)	EI (corr. not shown)	82	117	5020.0	0.575	1.000	ns
TG	value	EI (corr. shown)	FIC	82	201	8226.0	0.981	1.000	ns
TG	value	EI (corr. shown)	FIN	82	200	8435.5	0.702	1.000	ns
TG	value	EI (corr. not shown)	FIC	117	201	11214.0	0.486	1.000	ns
TG	value	EI (corr. not shown)	FIN	117	200	11512.5	0.809	1.000	ns
TG	value	FIC	FIN	201	200	20698.0	0.601	1.000	ns

Table A8: Pairwise Wilcoxon test for Voronezh decisions only

name	.y.	group1	group2	n1	n2	statistic	p	p.adj	p.adj.signif
CG	value	EI (corr. shown)	EI (corr. not shown)	82	117	5078.0	0.481	1.000	ns
CG	value	EI (corr. shown)	FIC	82	201	8333.5	0.882	1.000	ns
CG	value	EI (corr. shown)	FIN	82	200	8339.5	0.822	1.000	ns
CG	value	EI (corr. not shown)	FIC	117	201	11198.5	0.476	1.000	ns
CG	value	EI (corr. not shown)	FIN	117	200	11217.5	0.538	1.000	ns
CG	value	FIC	FIN	201	200	20200.5	0.931	1.000	ns
TG	value	EI (corr. shown)	EI (corr. not shown)	82	117	3682.0	0.005	0.029	*
TG	value	EI (corr. shown)	FIC	82	201	7304.0	0.129	0.387	ns
TG	value	EI (corr. shown)	FIN	82	200	6732.0	0.017	0.085	ns
TG	value	EI (corr. not shown)	FIC	117	201	13162.5	0.071	0.283	ns
TG	value	EI (corr. not shown)	FIN	117	200	12123.5	0.586	0.586	ns
TG	value	FIC	FIN	201	200	18682.5	0.215	0.430	ns

A.8 Shapiro–Wilk normality tests of main DVs

estimate	group1	group2	n1	n2	statistic	p	conf.low	conf.high	p.adj	p.adj.signif
-3.26e-05	EI (corr. shown)	EI (corr. not shown)	82	117	3776.0	0.007	-9.9999692	-0.0000491	0.042	*
-2.29e-05	EI (corr. shown)	FIC	82	201	7578.0	0.268	-4.0000548	0.0000249	0.536	ns
-3.64e-05	EI (corr. shown)	FIN	82	200	6646.5	0.007	-5.0000513	-0.0000338	0.042	*
6.30e-06	EI (corr. not shown)	FIC	117	201	13249.0	0.047	-0.0000635	4.9999776	0.188	ns
3.50e-05	EI (corr. not shown)	FIN	117	200	12024.5	0.655	-0.0000167	0.0000705	0.655	ns
-1.75e-05	FIC	FIN	201	200	18127.0	0.070	-0.0000377	0.0000279	0.209	ns

Table A9: Pairwise two sample Wilcoxon tests for difference in trust transfers

estimate	group1	group2	n1	n2	statistic	p	conf.low	conf.high	p.adj	p.adj.signif
5.0000336	EI (corr. shown)	EI (corr. not shown)	82	117	6198.0	0.000367	0.9999755	9.9999911	0.002000	**
0.0000341	EI (corr. shown)	FIC	82	201	8848.5	0.320000	-0.0000012	4.9999930	0.640000	ns
4.9999422	EI (corr. shown)	FIN	82	200	10512.5	0.000121	0.0000303	9.0000480	0.000726	***
-3.0000059	EI (corr. not shown)	FIC	117	201	9174.0	0.000773	-5.0000216	-0.0000307	0.002000	**
-0.0000517	EI (corr. not shown)	FIN	117	200	11208.5	0.516000	-0.0000585	0.0000423	0.640000	ns
0.0000770	FIC	FIN	201	200	24016.0	0.000443	0.0000539	4.9999504	0.002000	**

Table A10: Pairwise two sample Wilcoxon tests for difference in beliefs

B Ethics and competing interests statements

The authors have no competing interests to declare.

The experimental design was evaluated and approved by approved by German association for Experimental Economic Research, GfeW e.V., certificate number `bwcw68Gx`, available at https://gfew.de/ethik/bwcw68Gx.

References

Zafer Akın. Dishonesty, social information, and sorting. *Journal of Behavioral and Experimental Economics*, 80:199–210, 2019. ISSN 2214-8043. doi: 10.1016/j.socec.2019.03.009. URL https://www.sciencedirect.com/science/article/pii/S2214804318300466.

Staffan Andersson and Paul M. Heywood. The politics of perception: use and abuse of Transparency International's approach to measuring corruption. *Political studies*, 57(4): 746–767, 2009.

Omar Azfar and William Robert Nelson. Transparency, wages, and the separation of powers: An experimental analysis of corruption. *Public Choice*, 130(3):471–493, March 2007. ISSN 1573-7101. doi: 10.1007/s11127-006-9101-5. URL https://doi.org/10.1007/s11127-006-9101-5.

Ritwik Banerjee. Corruption, norm violation and decay in social capital. *Journal of Public Economics*, 137:14–27, 2016. ISSN 0047-2727. doi: 10.1016/j.jpubeco.2016.03.007. URL https://www.sciencedirect.com/science/article/pii/S0047272716300135.

Abigail Barr and Danila Serra. Corruption and culture: An experimental analysis. *Journal of Public Economics*, 94(11):862–869, 2010. ISSN 0047-2727. doi: 10.1016/j.jpubeco.2010.07.006. URL https://www.sciencedirect.com/science/article/pii/S0047272710000927.

Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, reciprocity, and social history. *Games and economic behavior*, 10(1):122–142, 1995.

Cristina Bicchieri. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press, December 2005. ISBN 978-1-139-44714-0.

Wagner Hugo Bonat, Paulo Justiniano Ribeiro Jr, and Walmes Marques Zeviani. Likelihood analysis for a class of beta mixed models. *Journal of Applied Statistics*, 42(2):252–266, 2015. Publisher: Taylor & Francis.

Juergen Bracht and Nick Feltovich. Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game. *Journal of Public Economics*, 93(9):1036–1044, 2009. ISSN 0047-2727. doi: 10.1016/j.jpubeco.2009.06.004. URL https://www.sciencedirect.com/science/article/pii/S004727270900067X.

J. Brandts and G. Charness. The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, 14(3):375–398, 2011.

Alessandro Bucciol and Marco Piovesan. Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology*, 32(1):73–78, 2011.

Alessandro Bucciol, Fabio Landini, and Marco Piovesan. Unethical behavior in the field: Demographic characteristics and beliefs of the cheater. *Journal of Economic Behavior & Organization*, 93:248–257, September 2013. ISSN 0167-2681. doi: 10.1016/j.jebo.2013.03.018. URL https://www.sciencedirect.com/science/article/pii/S0167268113000632.

Dmitrii Butrin. Konturnaja karta korrupcii, June 2019. URL https://www.kommersant.ru/doc/3990138.

Philipp Chapkovski. Conducting interactive experiments on Toloka. *Journal of Behavioral and Experimental Finance*, 37:100790, March 2023. ISSN 2214-6350. doi: 10.1016/j.jbef.2023.100790. URL https://www.sciencedirect.com/science/article/pii/S2214635023000047.

Gary Charness, Uri Gneezy, and Brianna Halladay. Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131:141–150, 2016. ISSN 0167-2681. doi: 10.1016/j.jebo.2016.08.010. URL https://www.sciencedirect.com/science/article/pii/S0167268116301779.

Liza Charroin, Bernard Fortin, and Marie Claire Villeval. Peer effects, self-selection and dishonesty. *Journal of Economic Behavior & Organization*, 200:618–637, August 2022. ISSN 0167-2681. doi: 10.1016/j.jebo.2022.06.024. URL https://www.sciencedirect.com/science/article/pii/S0167268122002207.

Nicholas Charron, Lewis Dijkstra, and Victor Lapuente. Mapping the Regional Divide in Europe: A Measure for Assessing Quality of Government in 206 European Regions. *Social Indicators Research*, 122(2):315–346, 2015. ISSN 1573-0921. doi: 10.1007/s11205-014-0702-y. URL https://doi.org/10.1007/s11205-014-0702-y.

Nic Cheeseman and Caryn Peiffer. The unintended consequences of anti-corruption messaging in Nigeria: Why pessimists are always disappointed. *ACE Working Paper*, 24, 2020.

Daniel L. Chen, Martin Schonger, and Chris Wickens. oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016. URL http://www.sciencedirect.com/science/article/pii/S2214635016000101.

Arndt Christiane. *Development Centre Studies Uses and Abuses of Governance Indicators*. OECD Publishing, 2006.

Ana Corbacho, Daniel W. Gingerich, Virginia Oliveros, and Mauricio Ruiz-Vega. Corruption as a Self-Fulfilling Prophecy: Evidence from a Survey Experiment in Costa Rica. *American Journal of Political Science*, 2016. URL http://onlinelibrary.wiley.com/doi/10.1111/ajps.12244/pdf.

David L Dickinson and David M McEvoy. Further from the truth: The impact of moving from in-person to online settings on dishonest behavior. *Journal of Behavioral and Experimental Economics*, 90:101649, 2021. ISSN 2214-8043. doi: 10.1016/j.socec.2020.101649. URL https://www.sciencedirect.com/science/article/pii/S2214804320306923.

Anja Dieckmann, Veronika Grimm, Matthias Unfried, Verena Utikal, and Lorenzo Valmasoni. On trust in honesty and volunteering among Europeans: Cross-country evidence on perceptions and behavior. *European Economic Review*, 90:225–253, November 2016. ISSN 0014-2921. doi: 10.1016/j.euroecorev.2016.01.011. URL https://www.sciencedirect.com/science/article/pii/S001429211630040X.

Andreas Diekmann, Wojtek Przepiorka, and Heiko Rauhut. Lifting the veil of ignorance: An experiment on the contagiousness of norm violations. *Rationality and Society*, 27 (3):309–333, 2015. ISSN 1043-4631. doi: 10.1177/1043463115593109. URL https://doi.org/10.1177/1043463115593109.

Eugen Dimant, Gerben A. van Kleef, and Shaul Shalvi. Requiem for a Nudge: Framing effects in nudging honesty. *Journal of Economic Behavior & Organization*, 172:247–266, April 2020. ISSN 0167-2681. doi: 10.1016/j.jebo.2020.02.015. URL https://www.sciencedirect.com/science/article/pii/S0167268120300597.

Dilyan Donchev and Gergely Ujhelyi. What do corruption indices measure? *Economics & Politics*, 26(2):309–331, 2014.

Bin Dong, Uwe Dulleck, and Benno Torgler. Conditional corruption. *Journal of Economic Psychology*, 33(3):609–627, 2012. ISSN 0167-4870. doi: 10.1016/j.joep.2011.12.001. URL https://www.sciencedirect.com/science/article/pii/S0167487011001759.

Armin Falk and Christian Zehnder. A City-Wide Experiment on Trust Discrimination. *Journal of Public Economics*, (0):–, 2013. ISSN 0047-2727. doi: 10.1016/j.jpubeco.2013.01.005. URL http://www.sciencedirect.com/science/article/pii/S0047272713000182.

Hanming Fang and Andrea Moro. Theories of statistical discrimination and affirmative action: A survey. *Handbook of social economics*, 1:133–200, 2011.

Silvia Ferrari and Francisco Cribari-Neto. Beta regression for modelling rates and proportions. *Journal of applied statistics*, 31(7):799–815, 2004. Publisher: Taylor & Francis.

FOM. The State of Domestic Corruption in the Russian Federation. Technical report, Ministry of Economic Development, Moscow, 2011. URL https://old.economy.gov.ru/minec/resources/116f09004739f0c7a2a4eeb4415291f1/doklad_kor.pdf.

Glynis Gawn and Robert Innes. Do Lies Erode Trust? *International Economic Review*, 59(1):137–161, 2018. ISSN 1468-2354. doi: 10.1111/iere.12265. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/iere.12265. 9 citations (Crossref) [2023-01-30] _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/iere.12265.

David Hugh-Jones. Honesty, beliefs about honesty, and economic growth in 15 countries. *Journal of Economic Behavior & Organization*, 127:99–114, 2016.

Ting Jiang. Other-Regarding Preferences and Other-Regarding Cheating – Experimental Evidence from China, Italy, Japan and the Netherlands. SSRN Scholarly Paper ID 2558814, Social Science Research Network, Rochester, NY, December 2014. URL https://papers.ssrn.com/abstract=2558814.

Noel D. Johnson and Alexandra A. Mislin. Trust games: A meta-analysis. *Journal of economic psychology*, 32(5):865–889, 2011.

Erin L. Krupka and Roberto A. Weber. Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524, 2013. URL http://sci-hub.cc/http://onlinelibrary.wiley.com/doi/10.1111/jeea.12006/full.

Nils C. Köbis, Jan-Willem van Prooijen, Francesca Righetti, and Paul A. M. Van Lange. "Who Doesn't?"—The Impact of Descriptive Norms on Corruption. *PLOS ONE*, 10 (6):e0131830, 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0131830. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0131830. 56 citations (Crossref) [2023-01-30] Publisher: Public Library of Science.

Nils C. Köbis, Marleen Troost, Cyril O. Brandt, and Ivan Soraperra. Social norms of corruption in the field: social nudges on posters can help to reduce bribery. *Behavioural Public Policy*, 6(4):597–624, 2022.

Johann Graf Lambsdorff. The methodology of the corruption perceptions index 2007. *Internet Center for Corruption Research, http://www. icgg. org/corruption. cpi_2006. html. Accessed*, 12, 2007.

Shaomin Li and Jun (Judy) Wu. Why some countries thrive despite corruption: The role of trust in the corruption–efficiency relationship. *Review of International Political Economy*, 17(1):129–154, February 2010. ISSN 0969-2290. doi: 10.1080/09692290802577446. URL https://doi.org/10.1080/09692290802577446. 53 citations (Crossref) [2023-01-30] Publisher: Routledge _eprint: https://doi.org/10.1080/09692290802577446.

Heather Mann, Ximena Garcia-Rada, Lars Hornuf, Juan Tafurt, and Dan Ariely. Cut From the Same Cloth: Similarly Dishonest Individuals Across Countries. *Journal of Cross-Cultural Psychology*, 47(6):858–874, 2016. ISSN 0022-0221. doi: 10.1177/0022022116648211. URL https://doi.org/10.1177/0022022116648211.

Massimo Mastruzzi, D Kaufmann, and A Kraay. The worldwide governance indicators: Methodology and analytical issues. *World Bank Policy Research Working Paper*, (5430), 2010.

Rob McCarney, James Warner, Steve Iliffe, Robbert Van Haselen, Mark Griffin, and Peter Fisher. The Hawthorne Effect: a randomised, controlled trial. *BMC medical research methodology*, 7(1):1–8, 2007.

Claire Mouminoux and Jean-Louis Rullière. Are we more honest than others think we are? 2021. URL https://hal.archives-ouvertes.fr/hal-01999536.

Sarah Necker, Benoit Le Maux, and David Masclet. Are people conditionally honest? The effects of stakes and information about others' behavior. Kiel, Hamburg: ZBW - Leibniz Information Centre for Economics, 2020. URL https://www.econstor.eu/handle/10419/224578.

Asmus Leth Olsen, Frederik Hjorth, Nikolaj Harmon, and Sebastian Barfort. Behavioral dishonesty in the public sector. *Journal of Public Administration Research and Theory*, 29(4):572–590, 2019.

Caryn Peiffer and Linda Alvarez. Who Will Be the "Principled-Principals"? Perceptions of Corruption and Willingness to Engage in Anticorruption Activism. *Governance*, 29(3):351–369, 2016. ISSN 1468-0491. doi: 10.1111/gove.12172. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/gove.12172.

Heiko Rauhut. Beliefs about Lying and Spreading of Dishonesty: Undetected Lies and Their Constructive and Destructive Social Dynamics in Dice Experiments. *PLOS ONE*, 8(11):e77878, 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0077878. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0077878.

Ritva Reinikka and Jakob Svensson. The power of information in public services: Evidence from education in Uganda. *Journal of Public Economics*, 95(7-8):956–966, 2011.

Innes Robert and Mitra Arnab. Is Dishonesty Contagious? *Economic Inquiry*, 51(1):722–734, 2013. ISSN 1465-7295. doi: 10.1111/j.1465-7295.2012.00470.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1465-7295.2012.00470.x. 52 citations (Crossref) [2023-01-30] _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1465-7295.2012.00470.x.

Anja Rohwer. Measuring corruption: a comparison between the transparency international's corruption perceptions index and the World Bank's worldwide governance indicators. *CESifo DICE Report*, 7(3):42–52, 2009.

Arthur Schram, Jin Di Zheng, and Tatyana Zhuravleva. Corruption: A cross-country comparison of contagion and conformism. *Journal of Economic Behavior & Organization*, 193:497–518, January 2022. ISSN 0167-2681. doi: 10.1016/j.jebo.2021.11.017. URL https://www.sciencedirect.com/science/article/pii/S016726812100490X.

Mariangela Sciandra and Irene Carola Spera. A model-based approach to Spotify data analysis: a Beta GLMM. *Journal of Applied Statistics*, 49(1):214–229, 2022. Publisher: Taylor & Francis.

Robert Slonim and Pablo Guillen. Gender selection discrimination: Evidence from a Trust game. *Journal of Economic Behavior & Organization*, 76(2):385–405, 2010. ISSN 0167-2681. doi: 10.1016/j.jebo.2010.06.016. URL https://www.sciencedirect.com/science/article/pii/S0167268110001368.

Giuliana Spadaro, Catherine Molho, Jan-Willem Van Prooijen, Angelo Romano, Cristina O. Mosso, and Paul A. M. Van Lange. Corrupt third parties undermine trust and prosocial behaviour between people. *Nature Human Behaviour*, pages 1–9, October 2022. ISSN 2397-3374. doi: 10.1038/s41562-022-01457-w. URL https://www.nature.com/articles/s41562-022-01457-w. 0 citations (Crossref) [2023-01-30] Publisher: Nature Publishing Group.

Eric M. Uslaner. Trust and Corruption. In Johan Graff Lamsdorf, Markus Taube, and Matthias Schramm, editors, *The New Institutional Economics of Corruption*, pages 76–92. Routledge, Loundon, 2004.

Danielle E. Warren and William S. Laufer. Are corruption indices a self-fulfilling prophecy? A social labeling perspective of corruption. *Journal of business ethics*, 88(4):841–849, 2009.

Rick K. Wilson and Catherine C. Eckel. Judging a Book by its Cover: Beauty and Expectations in the Trust Game. *Political Research Quarterly*, 59(2):189–202, 2006. ISSN 1065-9129. doi: 10.1177/106591290605900202. URL https://doi.org/10.1177/106591290605900202.

Jung-Yeop Woo and Uk Heo. Corruption and foreign direct investment attractiveness in Asia. *Asian Politics & Policy*, 1(2):223–238, 2009.

Nan Zhang. Changing a 'culture' of corruption: Evidence from an economic experiment in Italy. *Rationality and Society*, 27(4):387–413, 2015. ISSN 1043-4631. doi: 10.1177/1043463115605475. URL https://doi.org/10.1177/1043463115605475.

Daniel John Zizzo. Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98, March 2010. ISSN 1573-6938. doi: 10.1007/s10683-009-9230-z. URL https://doi.org/10.1007/s10683-009-9230-z.

Michael Zürn and Sascha Topolinski. When trust comes easy: Articulatory fluency increases transfers in the trust game. *Journal of Economic Psychology*, 61:74–86, 2017. ISSN 0167-4870. doi: 10.1016/j.joep.2017.02.016. URL https://www.sciencedirect.com/science/article/pii/S0167487016305931.