Unintended consequences of corruption indices: an experimental approach

Chapkovski, Philipp

30 March 2022

Unintended consequences of corruption indices: an experimental approach

This is a preliminary draft.
Please do not cite without the permission of the author.

Philipp Chapkovski*

March 29, 2022

Abstract

Using the results of a large-scale (N=900) online experiment, this paper investigates how the information about a group corruption level may harm intergroup relations. Corruption indices are widely used as a measure of quality of governance. But in addition to be a valuable tool for investors and policy makers for making informed decisions, they may also result in statistical discrimination: people from a more 'corrupt' region may be perceived as less trustworthy or more inclined to dishonest behavior.

We manipulated the amount of information people have about three different Russian regions in two simple behavioral games ('Cheating game' and Trust game). In a Cheating game after the main stage where they report whether they observed a head or a tail on a flipped coin, they guessed how many participants in each of the three regions reported more personally profitable outcome (heads) as well as make trasfer decisions in a standard trust game. In the baseline treatment we provided them with a set of generic information about each region (such as population size), while in the main treatment they were additionally informed about the degree of perceived corruption in each region. The presence of corruption information made people substantially overestimate the degree of dishonesty in more 'corrupt' regions and decreased the trust towards a person from this region. The results demonstrate the potentially harmful unintended consequences of corruption indices that have to be taken into account by policy makers.

Keywords: Corruption, experiments,online research, survey research

*Philipp Chapkovski (fchapkovskiy@hse.ru) is a Research Fellow at the International Laboratory for Experimental and Behavioural Economics at the Higher School of Economics in Moscow.

1 Introduction

In 2019 Russian Government approved the "Methodology for conducting sociological research to assess the level of corruption in the subjects of the Russian Federation" as a part of their National Anti-Corruption Plan. This document announced that the state will fund annual country-wide survey to measure perceived corruption level in each region [Butrin, 2019] (due to the forthcoming COVID-19 pandemics these plans were put on hold). Its aim was to create indices of 'petty' (domestic) and 'business' corruption which would allow to rank each Russian region along this dimension. The level of regional corruption was going to be based on a composite index of the set of questions about regularity of corrupt encounters and requests, the size of the bribes and the readiness of respondents to pay such bribes. The index was modelled after the regional corruption index composed by the Ministry of Economic Development[1]. This report showed a large regional variability in domestic corruption: varying from 15 points out of 100 in Jewish Autonomous Oblast to 80 in Krasnodar region.

The logic that motivates the state for creation of such indices is clear: they can be used as key performance indicators (KPI) for measuring efficiency of regional authorities, which is particularly helpful for such highly centralized state as Russia. Various indices that measure corruption on the country level were crucial for bringing the focus on wide-spread corruption, helping to indicate its main causes and paving the way out of it. Such well-known indices as Corruption Perception Index, composed by Transparency international or World Bank Governance indicators [Rohwer, 2009] were proved to be efficient policy tools to navigate investors' decisions and to create a pressure for reform on government and public.

However using these indices may also create a corruption trap, a vicious circle where international decision makers may drive the development aid out of more corrupt countries, making implementation of anti-corruption reforms impossible [Andersson and Heywood, 2009]. Furthermore, in addition to a direct potential harm, such indices limit a state ability to improve its governance indirectly by hitting the foreign investment [Woo and Heo, 2009].

These negative by-products of corruption indices mainly focus on the decision makers, such as state officials or large international investors. However many scholars have pointed out that these tools may create a kind of observer's or Hawthorne effect [McCarney et al., 2007], when the mere existence of such indices may change the environment they aim to objectively measure. For instance they send a signal to the population that corruption

[1] The original report[FOM, 2011] is available in Russian at: `https://old.economy.gov.ru/minec/resources/116f09004739f0c7a2a4eeb4415291f1/doklad_kor.pdf`

is something that is within the normatively allowed behavior. Indeed the large scale survey experiment in Costa-Rica has demonstrated that when people are aware about the growing number of co-citizens admitting that they have corruption encounters in the last 12 months, they are more eager to reveal their own similar experience [Corbacho et al., 2016]. The dependency between willingness to engage in corruption and the perceived corruption of the social surrounding has been confirmed using the European Values and the World Values surveys [Dong et al., 2012]. Knowledge about high degree of corruption can also undermine the popular will to fight against corrupt bureaucracy: the analysis of the data from 71 countries in the Global Corruption Barometer survey by Transparency International has demonstrated that perceived high levels of corruption limit citizens' willingness to actively oppose it [Peiffer and Alvarez, 2016].

Information about overall corruption level changes normative and empirical expectations about corrupt deals. Many people refuse to pay a bribe because they have intrinsic moral stance against it. But as most other norms this anti-corruption norm is susceptible to social pressure: it is shaped by a person's beliefs on how many others would give a bribe in a similar situation (empirical expectations) and what they think others believe a right thing to do (normative expectations) [Bicchieri, 2005]. The exposure to information that the society they live in is more corrupt than they think, shifts both normative and empirical expectations, making moral costs of such an action lower [Cheeseman and Peiffer, 2020].

In this paper I focus on one more unintended consequence of corruption indices and shifted expectations. An increasing public awareness of the perceived level of corruption may change people's beliefs regarding people from more corrupt regions, undermining trust towards them, and making others believe that they are more inclined to participate in morally dubious activity. In other words the corruption indices can trigger statistical discrimination[Fang and Moro, 2011], boosting prejudices towards people from specific regions.

Although most of the studies which focus on consequences of information about corruption, do it using country-level indices [Donchev and Ujhelyi, 2014, Warren and Laufer, 2009], we intentionally choose regional measures of corruption. Here we proceed from the premise that the regional stereotyping is markedly lower than the cross-country one. Furthermore by using only regional variation within one country we keep the country factor fixed. If those who trust are from the same country as those to be trusted, we control for any cultural prejudices towards different countries, including pre-existing notions of their degree of corruption, honesty and trustworthiness. Additionally this paper can contribute to the otherwise under-studied subfield of variation and perception of regional corruption which has been rarely in a focus of research due to the lack of data (with a notable exception of Charron et al. [2015]).

This paper puts together two strands of literature: the one that analyses unintended consequences of information about corruption and an extensive set of studies examining factors that shape human honesty, inclination for cheating and trust. Specifically we aim to track how knowledge about regional level of corruption push people to re-estimate their beliefs about honesty of participants from a specific region.

Similar to corruption, cheating, or vice versa honesty, is a social norm, being conditional on what others do (or what actor believes they would do). In other words, individual propensity to lie is defined to a certain extent by the share of others who lie in similar circumstances. Despite the vastness of literature on this subject, the results about the direction and size of the effect remain inconclusive. Some have found a positive effect (people cheat more observing others cheating) [Diekmann et al., 2015, Necker et al., 2020], others found that the link between beliefs and cheating is complicated: the beliefs about cheating *per se* do not make people more inclined for cheating, but the exposure to the actual cheating behavior, press people to adjust their behavior towards the majority [Rauhut, 2013].

To the best of my knowledge that is the first study that investigates a potential causal link between information about corruption and beliefs about others' honesty. But the relations between corruption and honesty have been studied before. It has been found that country-level corruption scores are significantly correlated with beliefs about dishonesty levels: in one of the randomized trials participants expected higher levels of dishonesty expected in more corrupt countries although their actual behavior did not correlate with their country corruption level [Mann et al., 2016]. In another study individual cheating behavior was observed more often in countries ranked as more corrupt in Corruption Perception Index (TI CPI) [Jiang, 2014]. The cross-country study involving public officials in 10 different countries has found that country-level indicators of corruption are strongly correlated with the average behavioral dishonesty [Olsen et al., 2019].

By manipulating an access to regional corruption index we were able to track whether this information changes individual estimates of the people who live in less and more 'corrupt' regions. We were also able to track the effect of corruption information on trust towards participants from these regions.

2 Methods

Main objective of this study is to measure an effect of information about regional corruption on estimates of honesty of a partner and trust towards him/her. We operationalize beliefs about honesty by eliciting participants' estimates of the share of other participants who would report that they had observed a 'head' when they flipped a coin in a standard 'cheating' game [Bucciol and Piovesan, 2011]. The belief elicitation was incentivized: each correct guess by a participant increased their payoff [Krupka and Weber, 2013].

The trusting behavior was operationalized through participants' transfers as first movers in a Trust game. In a standard Trust or Investment game [Berg et al., 1995] participants are matched into groups of two. Both players receive a certain initial amount (endowment) and are assigned to one of two non-symmetrical roles. A first mover (*Trustor*) can choose to send any share of the endowment to the Second mover (*Trustee*). This amount is tripled by the experimenter. A second mover can send back any amount out of the multiplied sum received. Thus the amount sent by a first mover reflects his/her belief that a second mover will at least partially reciprocate his trust, while an amount returned by a second mover reflects his/her actual trustworthiness. A strictly profit-maximising strategy would dictate to a second mover to send nothing back from the multiplied amount received from Trustor. Since the trust is expected not to be reciprocated, a rational Trustor transfers zero to a Trustee. An empirical data based on dozens of published lab experiments with Trust game does not confirm this profit-maximising intuition: on average Trustors send 50.2% of their original endowment, while Trusteers return 37.2% back [Johnson and Mislin, 2011].

Across all treatments respondents participated both in belief elicitation and trust game stages. We check for the order effect by randomizing the 'cheating' and 'trusting' stages of the game and taking this order in account in regression models.

To measure an effect of information about a partner's region on the behavior of an actor we split the population into two unequal parts (see Figure 1). The first part was recruited exclusively from Moscow region, the second part was recruited from Moscow, Arkhangelsk and Voronezh regions (they are marked in Figure 1 as subindices *m, a* and *v* correspondingly). First pool of participants engaged in two stages (Cheating game, *CG* and Trust game, *TG* in a random order). After the CG stage they also reported their beliefs (designated as BE_i arrows on Figure 1) about the share of participants in each region from the second part of subpopulation who will report 'head' in a *CG* stage. In the stage 2, they also made three decisions in a role of first movers in Trust game (designated as TS_i on Figure 1). The second pool participated in CG stage followed by TG stage in a role of second movers (without randomization). The number of participants recruited for each treatment for the first subpopulation and for each region in the second subpopulation is reported in table 1.

These specific regions were selected as targets for evaluation of honesty and trust for several reasons. The main reason is that they are located at the top, middle and bottom of the ranking of the regions according to the index of domestic corruption [FOM, 2011]: see the index values for each region in rightmost column of Table 1. The corruption index that serves as a main intervention tool in this study is based on the report by the Ministry of Economic Development of 2011. The data used to compose the index of regional corruption was collected by Public Opinion Foundation, one of the largest Russian pollsters. The survey was conducted in 70 Russian regions ($n = 17.500$), representing 94.5% of the population. The index is composed by summing up frequency of responses in each region along four dimensions (the risk of being asked for a bribe, share of those who ever paid a bribe, a readiness to give a bribe and an average size of a bribe given). The index per se is not relevant for the purpose of this study because the main aim of it is to see whether provision of such aggregated information changes the behavior. What is crucial for us though is the fact that Russia shows a large heterogeneity of perceived corruption level across regions (see map of Russian regions from this report in Appendix, Figure A2) where Arkhangelsk region has 4 times lower index value than more 'corrupt' Voronezh region.

Some decisions participants had to make and the payoffs defined by them are by definition interdependent. We asked participants from the first pool to guess how many participants from the second pool in each region would report 'head' in a cheating game. It is possible to evaluate how close their guesses were to the true value only after the second pool of participants reported their decision in a Cheating game. Similarly, first movers in a trust game can receive their payoffs for the TG stage only after second movers make their decisions about returning transfers. For solving this issue we made the decisions asynchronous using the strategy method [Brandts and Charness, 2011]. Specifically, when we elicited beliefs after CG stage we informed participants that they have to estimate share of those who reported 'heads' in each of the region that participate in the study. After collecting CG decisions from the second pool, we matched first pool beliefs about second pool decisions in each region and calculated their payoffs. In Trust game first movers were informed that they can be matched to a second mover from the second pool, who can be from one of three regions, and thus they have to make the transfer decision three times, but only one of these decisions will be implemented based on the region from where the second mover is. The second movers were asked to decide what share of the transferred and multiplied amount they would like to send back to a first mover. This unconventional decision was made because second pool participants were not subject of

First movers in TG		Second movers in TG		
Treatment	N	Region	N	Corruption index
FIC	200	Moscow	100	17
FIN	200	Arkhangelsk	100	58
EI	200	Voronezh	100	81

Table 1: Distribution of participants across treatments and regions

the treatment manipulation and their decisions are not used in the analysis. After both first and second pool decisions were collected, we matched transfers of first and second movers and calculated corresponding payoffs. Unlike the *CG* stage beliefs that were paid based on the overall values of second pool participants, in *TG* stage we matched two first movers from the first pool with one second mover from the second pool. That was dictated by smaller available pools from Arkhanglesk and Voronezh regions, where we could not guarantee to have more than 100 participants in short time.

The perception of a stranger from a specific region may be shaped not only by information about this region in general, but also by two additional factors: it can be a region of origin for a person making evaluations, and he or she can be personally connected to the region through friends, relatives, experience and some idiosyncratic knowledge (i.e. having a personal conflict with someone from this region). To check whether the corruption information about one's own region can change the estimates of honesty and trusting behavior we also included Moscow. Since the only difference across treatments is the presence or a lack of regional corruption index, the difference in honesty estimates and trusting behavior toward their own region among Moscovites could only be explained by this manipulation.

2.1 Treatments

Participants from the first pool were assigned randomly to one of three treatments: {*FIC*, *FIN* and *EI* } that differed only by the amount of information provided about each region that participated in the second pool of the study.

Fixed Neutral information,*FIN* In this treatment participants were able to see the following information about each region (in random order): Gross Regional Product (*GRP*) per capita; Average population age (*pop_age*), and Consumer Price Index (*CPI*). The data was taken from Federal State Statistics Service regional database for 2020.

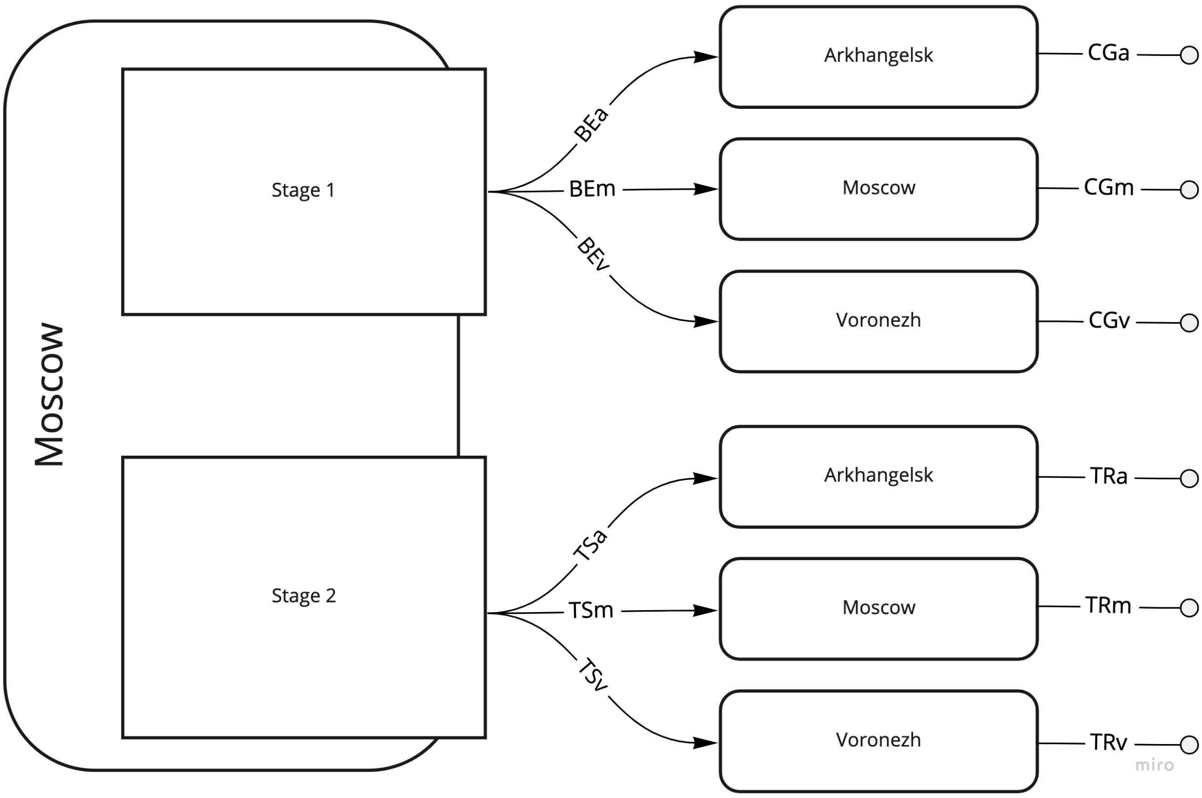

Figure 1: Gameflow across regions. Legend: *BE*: Belief elicitation; *CG*: Cheating game reporting decision; *TS*: First mover transfer in Trust game; *TR*: Second mover transfer in Trust game. Cities legend: *a*: Arkhangelsk; *m*: Moscow; *v*: Voronezh

Fixed information + Corruption, *FIC* In this treatment in addition to three indicators of *FIN* treatment the perceived corruption index was reported. We used the only available regional corruption index provided in the Ministry of Economic Development report of 2011 [FOM, 2011].

Endogenous choice of information, *EI* Since the presence of the regional corruption index may affect people's estimates not only directly but via experimenter demand [Zizzo, 2010], in *EI* treatment participants could choose 3 indicators out of 6 available. In addition to 4 indicators listed above (GRP, pop_age, CPI and corruption), we also let them choose from unemployment level and the birth rate per 1,000 inhabitants. Thus in *EI* treatment the regional information screen was preceded by the stage where participants could choose the indicators they would like to see. The order of these indicators was randomized as well.

Below we present the game flow for the order where CG stage comes first.

Cheating game, *CG* : In the standard coin flipping game [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 one of the numerous online services for flipping a virtual coin. If they reported the head, their payoff was increased by $1, if the tail was reported the payoff remained unchanged.

Information about regions revealed, *RI* If *CG* stage comes first, then participants were provided with a three-column table with indicators describing each region. The order in which regions were shown as well as the order in which indicators were listed for each region were randomized for each participant. The example of information available to participants in *FIC* treatment is provided in Figure 2.

Cheating game - Belief elicitation, *BE* In this stage *CG* we elicited their beliefs about the share of others who report head. For each of three regions participating in the study they gave their estimates on how many out of 100 participants would report "head". For each correct guess (within a margin of error of 10 percentage points) they could earn an extra bonus of $1. They could see the information about each region while making the estimates (For an example of the *BE* stage for *FIC* treatment see screenshot at Figure 2).

Trust game - First move, *TG* After participants finished the belief elicitation stage, they were informed that the second part of the study begins. The instructions for the standard Trust game [Berg et al., 1995] were shown, and they had to pass the comprehension test. After that their role was announced. Participants of the first pool were assigned to the role of a first mover (neutrally named *Participant A*), and participants of the second pool were assigned to the role of a second mover (*Participant B*). Participants A were told that their decisions will be matched with a decision of a participant who can be from one of the three regions, thus they have to make the transfer decision for each region but only one of them will be relevant. They could see the same information for each region that they were able to see during *BE* part, with the same order of indicators and regions. An example of the TG decision stage for *FIC* treatment is shown at the Figure 3.

As it can be seen from this game flow (and also from the flowchart in Figure A1) the participants were not aware of the content of the second part when they started the first part of the study. They were informed however in the beginning of the study that in both parts they can increase their final bonus, but only one of two parts will be randomly chosen for payment. This randomization of payments is a standard procedure in order to avoid hedging and wealth effect [Charness et al., 2016]. Likewise while making their decisions about reporting the results of coin flipping in CG stage they were not aware about the forthcoming incentivized belief elicitation (BE) stage. That was done to guarantee that their potentially 'cheating' decision will not be affected by the anticipated BE stage. The opposite is not true: obviously the reporting decision in CG influenced their beliefs (see section 3). But since the main research focus of the paper is on the effect of information on their beliefs and trusting behavior, and the experience of reporting in CG made participants better understand what kind of choices faced other participants, we made a decision to introduce belief elicitation part after CG and further control for their own decision in *CG* when we check for their beliefs.

Before running the study we preregistered[2] four hypotheses.

H1 *With additional information about the regional level of corruption, people provide lower honesty estimates for people from more corrupt regions.* So we do expect that in the belief elicitation stage of *FIC* treatment participants will evaluate number of heads reported in more corrupt region (specifically Voronezh) higher than number of heads reported in less corrupt region (Arkhangelsk) while this gap will be smaller in the *FIN* treatment where no information about corruption is provided. Likewise, when information is chosen by participants themselves, in *EI* treatment, we will observe a larger gap in honesty evaluations by those who choose to obtain corruption information than by those who preferred other indicators.

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

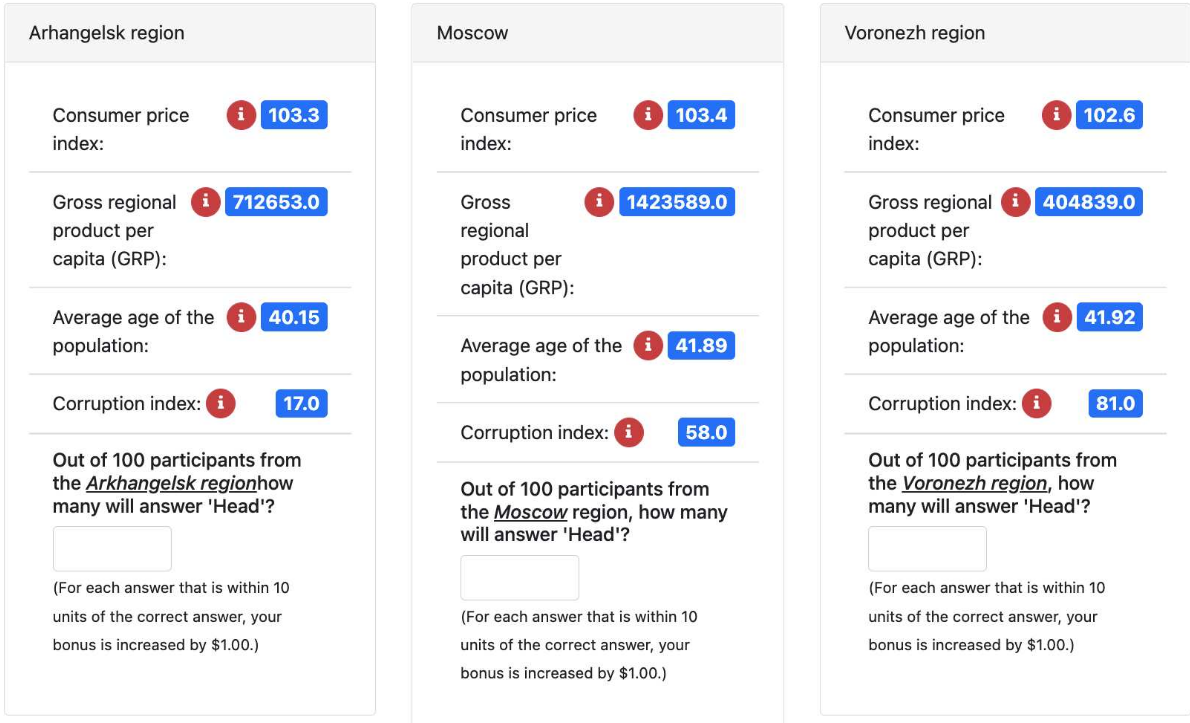

Figure 2: Screenshot of the decision stage of CG game, *FIC* treatment (automatic translation from Russian)

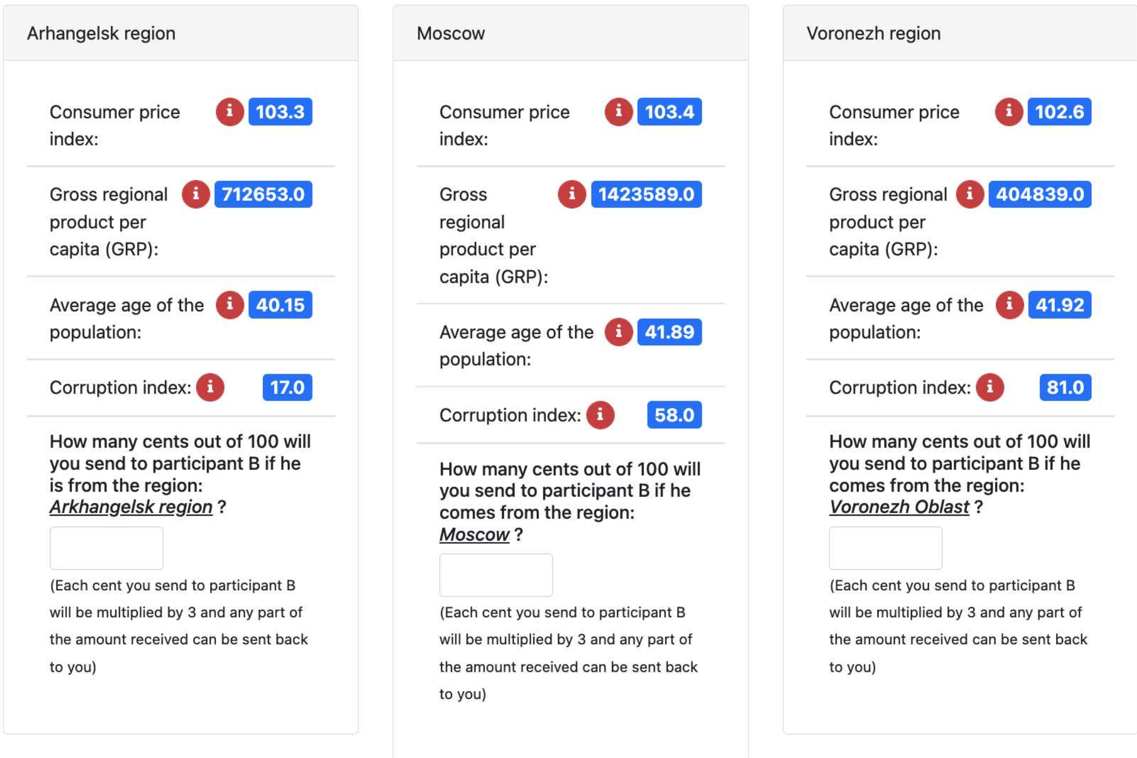

Figure 3: Screenshot of the decision stage of TG game, *FIC* treatment (automatic translation from Russian)

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 4: Questions regarding regional knowledge (automatic translation from Russian)

H2 *With additional information about the regional level of corruption, first movers in a trust game will make lower transfers towards second movers from more corrupt regions.* Similarly to *H1* we expect that participants who have (or chose to have as in *EI* treatment) the information about regional corruption will transfer more to potential partners from less corrupt regions, while this gap between regional transfers will be smaller in the *FIN* treatment.

H3 *Participants' estimates of honesty of people from their own region are different when we give them the information about the corruption level of their own region.* When participants evaluate the number of heads reported within their own region (Moscow), when they observe (*FIC* treatment) or choose to observe (*EI* treatment) regional corruption index of their own region they will estimate number of heads reported higher than in the case when this information is not available (*FIN* treatment) or they chose not to have (in *EI* treatment).

H4 *The transfers of first movers in a trust game towards people from their own region are different when they are additionally provided with the information about the corruption level of their own region.* Similarly to *H3* transfers to trustees from their own region (Moscow) are expected to be lower when the information of corruption in Moscow is available.

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.

3 Results

Here we present the results based on the decisions made by the first pool of participants (see 'Methods' section for details). There were three sessions, with planned number of participants of 200 in each one: each session corresponded to one of three treatments: *FIC*, *FIN* or *EI*. All sessions were run within the same day, on 17th of November 2021 on Toloka crowdsourcing platform.

Before we present the test results for each specific hypothesis separately, it makes sense to demonstrate an overall picture. For each participant we calculated a difference between their beliefs (*belief_diff* variable) and transfers (*trust_diff*) towards Voronezh (allegedly highly 'corrupt' city) and Arkhangelsk (low corrupt city) across three treatments. Thus $belief\_diff = cg\_belief_{Voronezh} - cg\_belief_{Arkhangesk}$ and $trust\_diff = tg\_decision_{Voronezh} - tg\_decision_{Arkhangesk}$. For the treatment with endogenous information choice (*EI*) we report results for those who chose corruption index (*EI (corr.shown)*) and those who didn't. On average when the corruption information was available participants believed that share of those reported heads in the more corrupt region would be higher by 6.57 and 5.02 points (for *FIC* and *EI-Corr* treatments), while this difference was roughly zero for no-information treatments (left panel of the Figure 5).

While making their first movers' transfer decisions, they transferred towards more 'corrupt' Voronezh 11.46 cents less for *EI* and 6.97 cents less in *FIC* treatment, while for no-information treatment the gap between transfers was not significantly different from 0 (0.76 and -2.12 for *EI-No corr* and *FIN* treatments correspondingly, right panel of the Figure 5).

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

H1 hypothesis is confirmed. As it can be seen in Figure 6, there is a difference in treatments with corruption information (*FIC* and *EI with corruption chosen*): 7.5% growth for *FIC* and 10% growth for *EI*, and there is no difference for treatments with no corruption information (either *FIN* or *EI with corruption not chosen*).

The pairwise Mann-Whitney-Wilcoxon tests checking for differences between beliefs about honesty of Arkhangelsk and Voronezh confirm the intuition that we can draw from the graph. In the Table 2 it is seen that the difference between beliefs about the

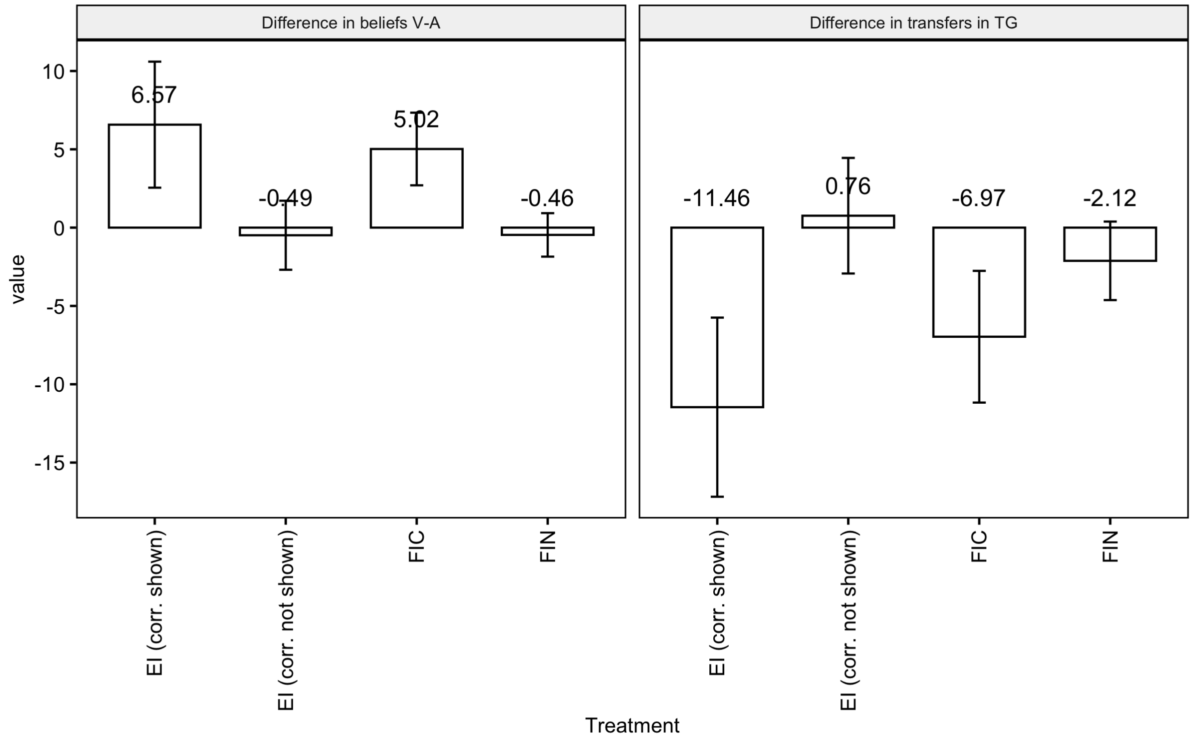

Figure 5: Difference in beliefs and trust within-subject

number of heads reported in high and low corruption entities are significant in *FIC* treatment and among those who chose to observe this indicator in *EI* treatment, and it is not significant in neutral (*FIN*) treatment and among those in *EI* who chose other than *corruption* indicator to observe for each region.

In table 3 we also report results of beta regressions where the dependent variable is a belief about the share of respondents who report 'head' in *CG* stage of the study. The baseline target city here is Arkhangelsk, and the baseline subtreatment is no-information treatment *FIN*. The three models presented in the table 3 only differ in degree of controls. The Model 1 is the most basic one, checking for interaction between target city and treatment without additional controls. Model 2 additionall controls for age, gender, order in which *CG* and *TG* stages were played, and for their own decision in *CG* stage. Finally Model 3 also controls for educational attainment, marital and employment status, income level and difference in knowledge regarding two region.

As it can be seen from Table 3, in all three models beliefs about Voronezh honesty in information treatments (*EI-Corr* and *FIC*) are substantially higher than in no-information treatment. The order of the game did not play any substantial role. The only factor that was significant apart from treatments was their own decision

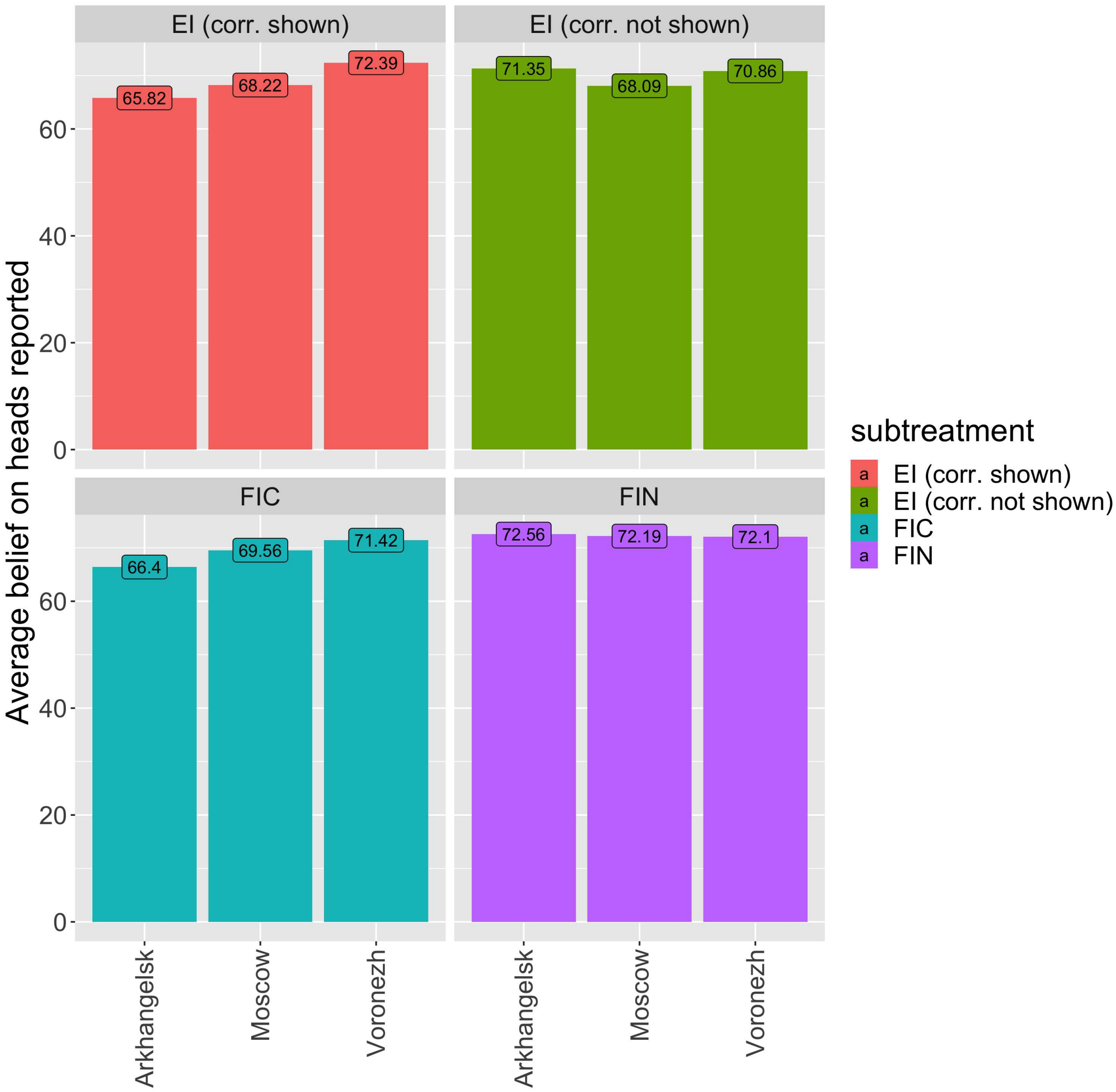

Figure 6: Average beliefs about cheating game outcomes

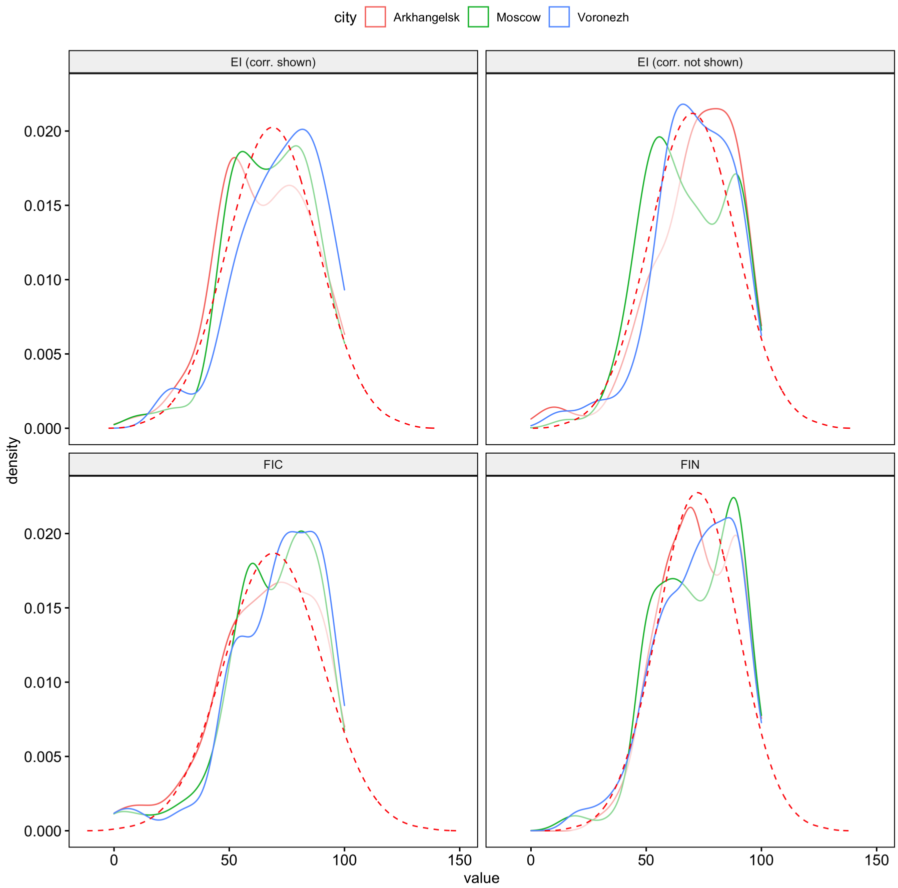

Figure 7: Distribution densities for honesty beliefs

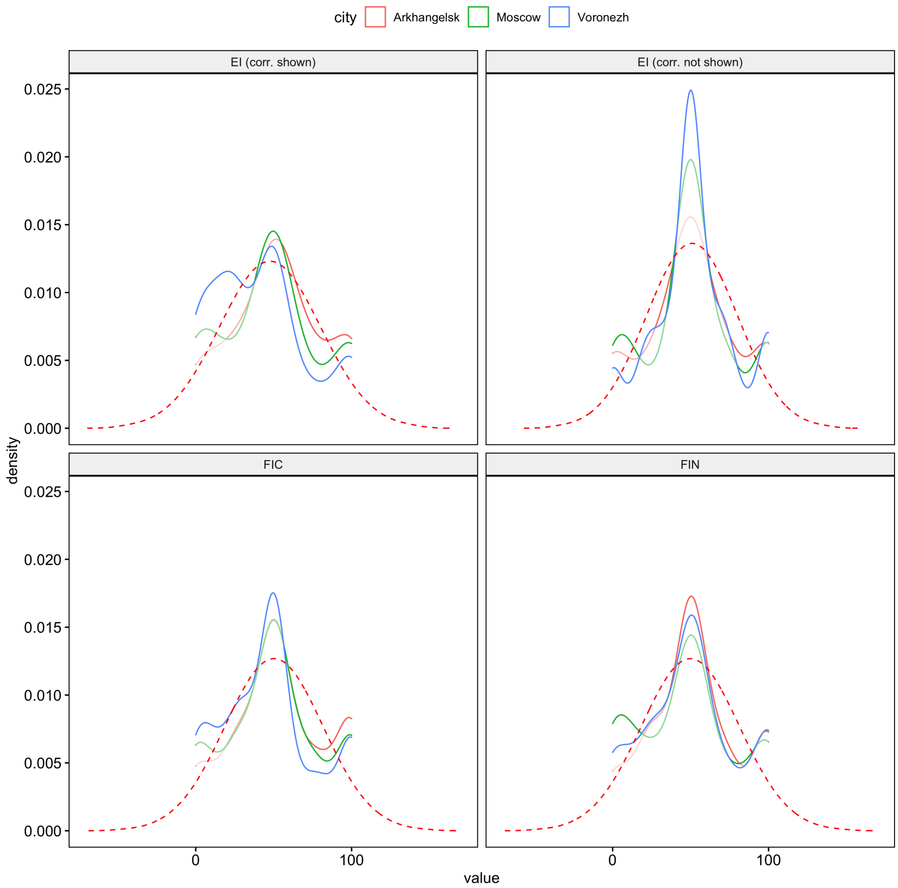

Figure 8: Distribution densities for trust beliefs

Table 2: pairwise Mann-Whitney-Wilcoxon tests on belief difference

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

in *CG* stage: those who reported head have significantly higher odds to provide higher estimates of others who report heads. That makes our finding similar to [Mouminoux and Rullière, 2021] where they have also found that people who cheat tend to have higher beliefs about dishonesty of others.

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

H2 hypothesis is also confirmed (see though some inconclusive results from beta regressions in Table 5). Very similar to the differences in honesty beliefs that we expect to see and indeed saw in *H1*, we saw the differences in transfer decisions made by first movers in a Trust game. In the *EI-Corr* treatment they transfered to Voronezh (more 'corrupt' city) 21% less than to potential second mover from Arkhangelsk. This difference was less visible when information was given, and not chosen, as in *FIC* treatment: 13% less towards Voronezh. No such difference were observed in no-information treatments (*FIN* and *EI-no corr*) - see Figure 9. The pairwise Mann-Whitney-Wilcoxon tests reported in table 4 confirm that this difference in transfers is significant for *FIC* and *EI-corr* and non significant for *FIN* and *EI-no corr*.

The results of beta regressions are shown in Table 5. They confirm that trust decisions towards Voronezh are smaller than towards Arkhangelsk in *EI-corr* treatment. For *FIC* treatment this difference is barely significant. This time in extended models (Models 2 and 3) where we control for game order, and reporting decisions in *CG* stage, these two factors are important: those who reported head trust less than those who reported tail.

H3 *Participants' estimates of 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 towards people from their own region are different when they are additionally provided with the information about the*

	Model 1	Model 2	Model 3
(Intercept)	1.073***	0.632***	0.540**
	(0.073)	(0.137)	(0.182)
targetvoronezh	−0.026	−0.026	−0.026
	(0.052)	(0.052)	(0.052)
subtreatmentEI (corr. not shown)	−0.074	−0.046	−0.056
	(0.119)	(0.118)	(0.117)
subtreatmentEI (corr. shown)	−0.310*	−0.330*	−0.332*
	(0.134)	(0.132)	(0.132)
subtreatmentFIC	−0.301**	−0.283**	−0.272**
	(0.102)	(0.101)	(0.101)
targetvoronezh × subtreatmentEI (corr. not shown)	0.005	0.006	0.005
	(0.087)	(0.087)	(0.087)
targetvoronezh × subtreatmentEI (corr. shown)	0.377***	0.379***	0.378***
	(0.097)	(0.097)	(0.097)
targetvoronezh × subtreatmentFIC	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_firstTRUE		0.010	0.000
		(0.077)	(0.077)
as.factor(cg_decision)1		0.394***	0.398***
		(0.083)	(0.083)
education			0.054
			(0.037)
marital			0.041
			(0.035)
employment			−0.024
			(0.020)
income			0.009
			(0.035)
knowledge_diff			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 3: Beta regressions, DV: honesty beliefs

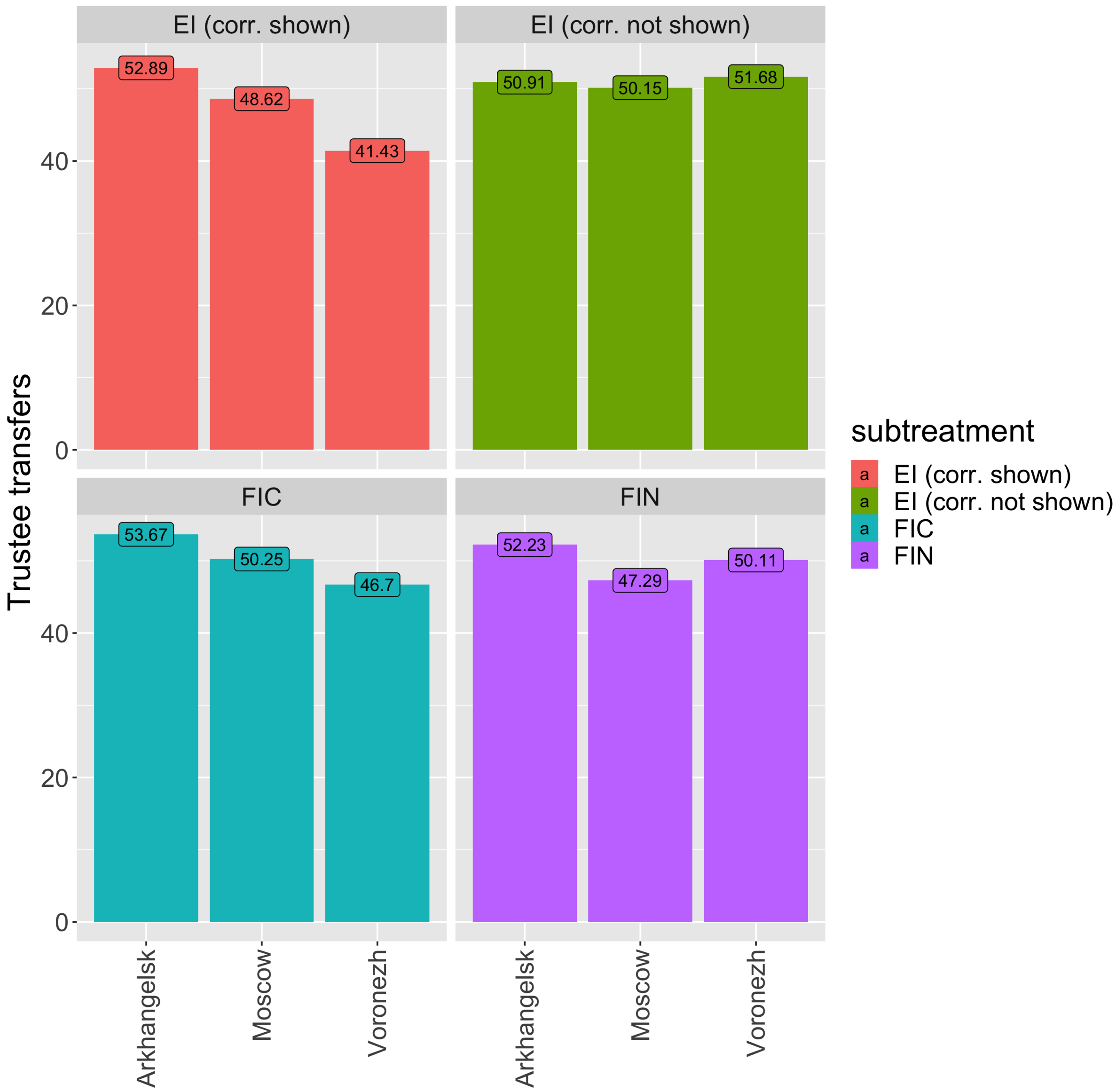

Figure 9: Average transfers by first movers in a Trust game

Table 4: pairwise Mann-Whitney-Wilcoxon tests on transfers in TD

subtreatment	group1	group2	n1	n2	statistic	p.adj	p.adj.signif
EI (corr. shown)	Arkhangelsk	Voronezh	82	82	4165.0	0.008	**
EI (corr. not shown)	Arkhangelsk	Voronezh	117	117	6770.0	0.885	ns
FIC	Arkhangelsk	Voronezh	201	201	22933.5	0.017	*
FIN	Arkhangelsk	Voronezh	200	200	20761.5	0.504	ns

	Model 1	Model 2	Model 3
(Intercept)	0.161	0.504*	0.668*
	(0.106)	(0.196)	(0.261)
targetVoronezh	−0.115	−0.115	−0.115
	(0.092)	(0.092)	(0.092)
subtreatmentEI (corr. not shown)	−0.121	−0.123	−0.122
	(0.173)	(0.172)	(0.172)
subtreatmentEI (corr. shown)	0.021	0.024	0.032
	(0.196)	(0.194)	(0.195)
subtreatmentFIC	0.074	0.050	0.047
	(0.149)	(0.149)	(0.149)
targetVoronezh × subtreatmentEI (corr. not shown)	0.183	0.181	0.180
	(0.153)	(0.153)	(0.153)
targetVoronezh × subtreatmentEI (corr. shown)	−0.434*	−0.432*	−0.430*
	(0.173)	(0.173)	(0.173)
targetVoronezh × subtreatmentFIC	−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_firstTRUE		0.200+	0.206+
		(0.109)	(0.109)
as.factor(cg_decision)1		−0.267*	−0.274*
		(0.118)	(0.119)
education			−0.023
			(0.053)
marital			0.005
			(0.050)
employment			0.000
			(0.028)
income			−0.040
			(0.050)
knowledge_diff			−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 5: Beta regressions, DV: trust trasfers in TG

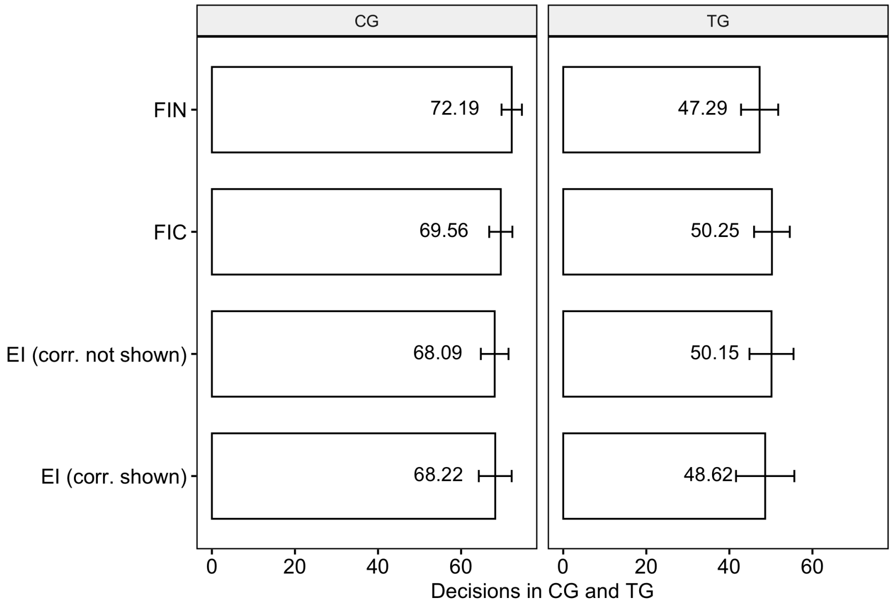

Figure 10: Overall decisions about Moscow in *CG* and *TG*

corruption level of their own region.

As the following analysis shows, neither *H3* nor *H4* hypotheses are confirmed by the existing data. We combine two hypotheses (*H3* and *H4*) together due to their similarity from the point of view of analysis. As the Figure 10 shows, there is no significant differences either in honesty beliefs or in trusting decisions across treatments. The non-parametric tests of differences in means for CG beliefs (Kruskall-Wallis test, KW statistic=5.72, p=0.126) or TG decisions (KW statistic=1.17, p=0.759) showed no difference. Pairwise Mann-Whitney-Wilcoxon test results are reported in Appendix (Table A2)

3.1 Endogenous information

In *EI* treatment participants were able to choose from six different indicators three that they are able to observe for each of the three regions that participated in the study. The corruption was chosen by 82 (41%) participants of 199 participating in *EI* treatment.

We also checked for two indicators that may potentially provide us some insights on why people choose to observe corruption index per region. These are whether they played *CG* or *TG* stages first (that may inform us for which of two dimensions: honesty or trust corruption can be more important), and whether those who reported head themselves, choose to observe corruption more often. Both indicators are shown in Figure **??**. It should be taken into account that the population sizes in subgroups within *EI* treatments

Table 6: Kruskal-Wallis tests for Voronezh and Arkhangelsk

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	

are relatively small (total size of the treatment is 200, number of people who chose to observe corruption index is 82), so no statistically significant results can be obtained from the available data.

Finally, we analysed differences in means of honesty beliefs and trusting decisions towards Arkhangelsk and Voronezh separately. That can give us some insights whether the observation of corruption index makes people re-estimate their attitudes towards more or less corrupt regions. Indeed, as results of Kruskal-Wallis tests show (Table **??**), there is difference in honesty beliefs across subtreatments for Arkhangelsk, and no difference for trusting decision. Vice versa, trusting decisions differ for Voronezh and are the same across treatments for beliefs. The additional pairwise tests and graphs for single cities are available in Appendix (Figures A4 and A3 and tables A3 and A4). These results may indicate that being exposed to corruption information, people start trusting less to those who live in more corrupt cities, but no more trusting occurs towards those who live in less corrupt cities. The situation is the opposite with honesty estimates: people re-estimate their evaluation of honesty about inhabitants of less corrupt regions, but this does not happen with their estimates of honesty of less corrupt regions.

4 Discussion

Wrapping up, we may summarize that information about corruption may corrupt human relations in quite unforeseen ways. As we saw, it undermines trust and make people believe that others are less honest than they would think if they were not aware of this information. The most worrisome interpretation of this result is that people are easily susceptible to the information manipulation that leads them to the statistical discrimination. Specifically they easily start drawing the conclusions about honesty and trustworthiness of a single person or a small subgroup based on an indicator relevant for the entire region.

This study is just a first attempt to approach an effect of corruption information on human behavior. Thus, it has some obvious limitations. We do not fully understand how exactly the mechanism of this statistical discrimination works. For instance we may

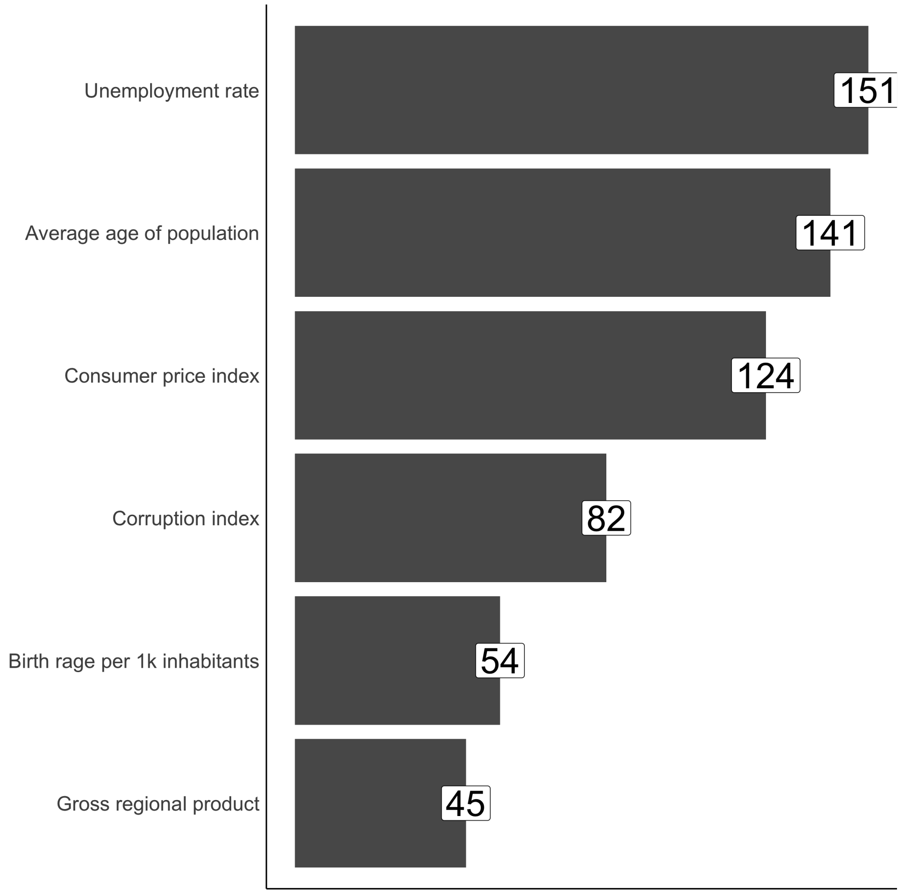

Figure 11: How often the indicators were chosen in *EI* treatment

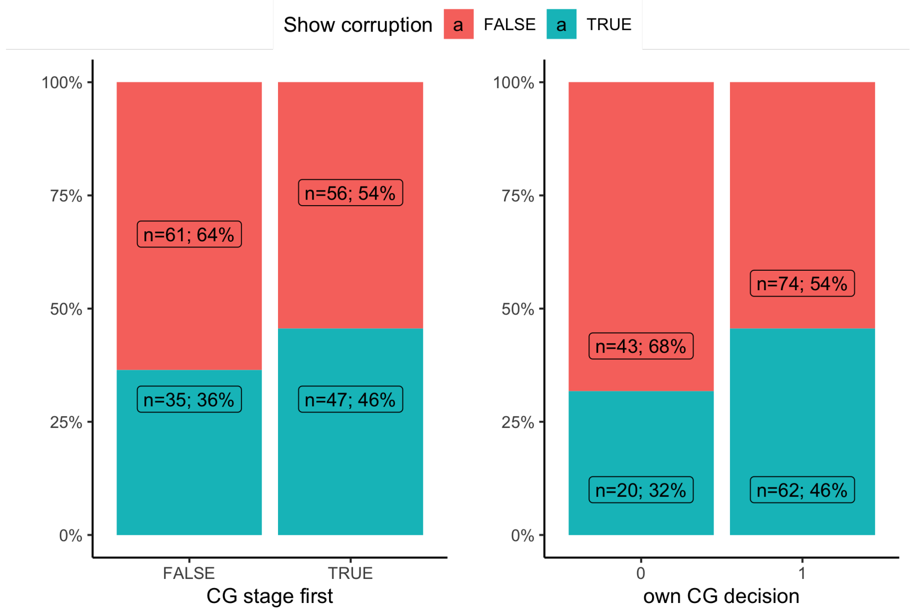

Figure 12: Decision to observe corruption by CG decision and game order

speculate that there is a certain degree of interaction between individual honesty and the effects of corruption information has both on trust and dishonesty, but the current design is limited in checking this hypothesis. It may happen that those who incline to cheat feel more trust to the similar people, so a certain degree of group identity may mitigate the adverse informational effect.

The share of actual cheating and beliefs about this cheating should be taken with caution because the experiment was conducted online and online audience has tendency to cheat more often than the one in the lab and be in general less trusting [Dickinson and McEvoy, 2021]. Furthermore, the external validity of the lab or online experiments especially in such sensitive topics as honesty and corruption should be taken with the grain of salt. Using existing techniques it would be important to double-check the conclusions of this paper in the field: for instance by investigating whether people are more reluctant to hire people when they know about the corruption level of their region of origin. We do believe that the current paper may be a first step towards more nuanced study of unintended consequences of corruption indices.

A Appendix

A.1 Flowchart of screens

Source: [FOM, 2011]

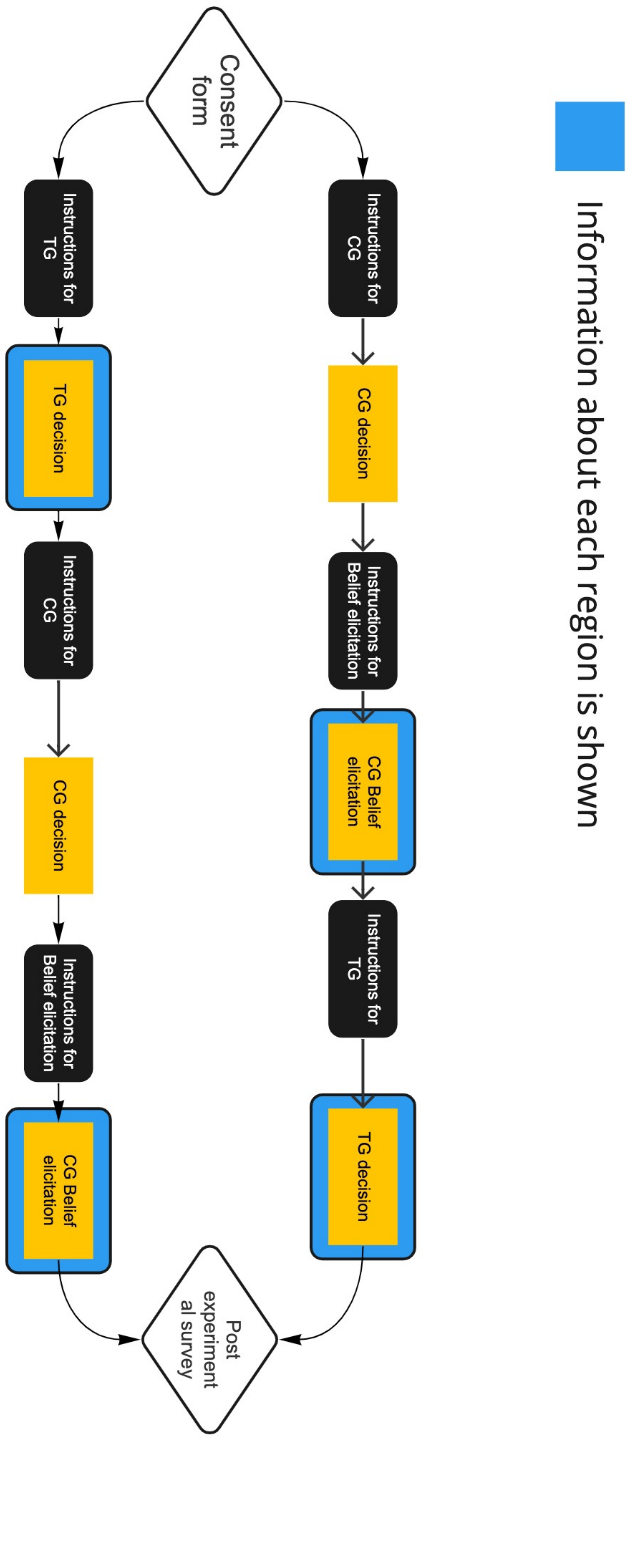

Figure A1: Flowchart of screens

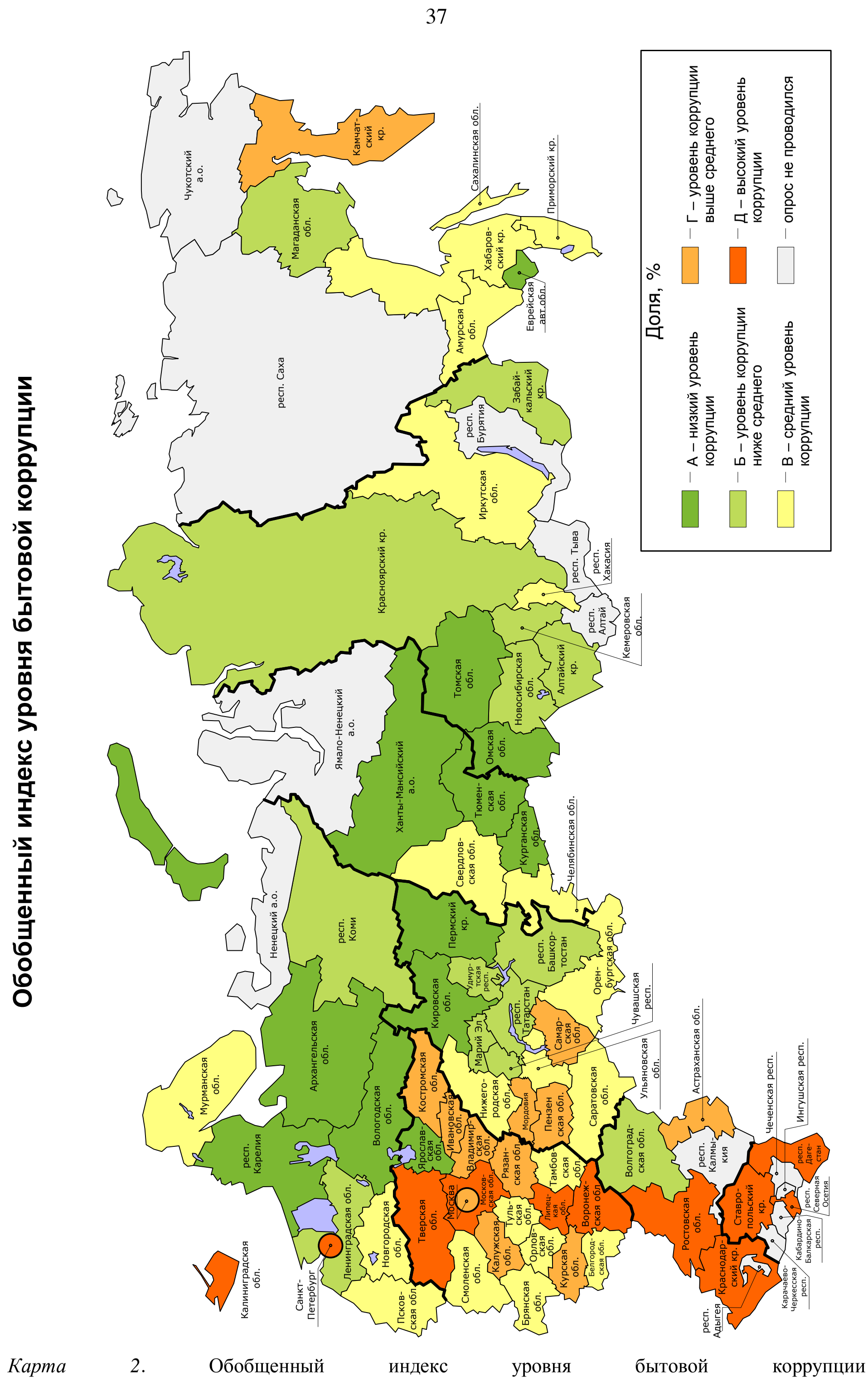

Figure A2: Flowchart of screens

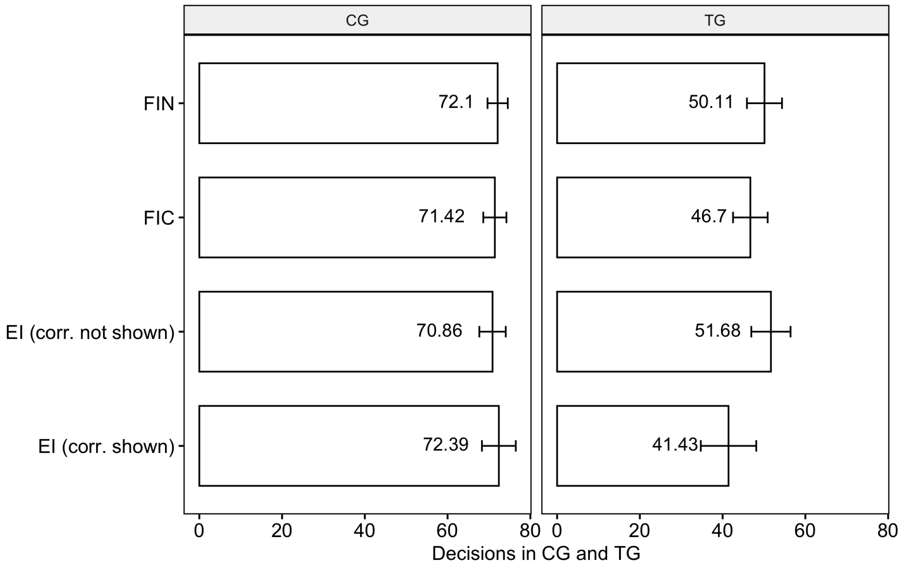

Figure A3: Overall decisions about Voronezh in *CG* and *TG*

A.2 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 A1: Results of Shapiro-Wilk Normality Tests

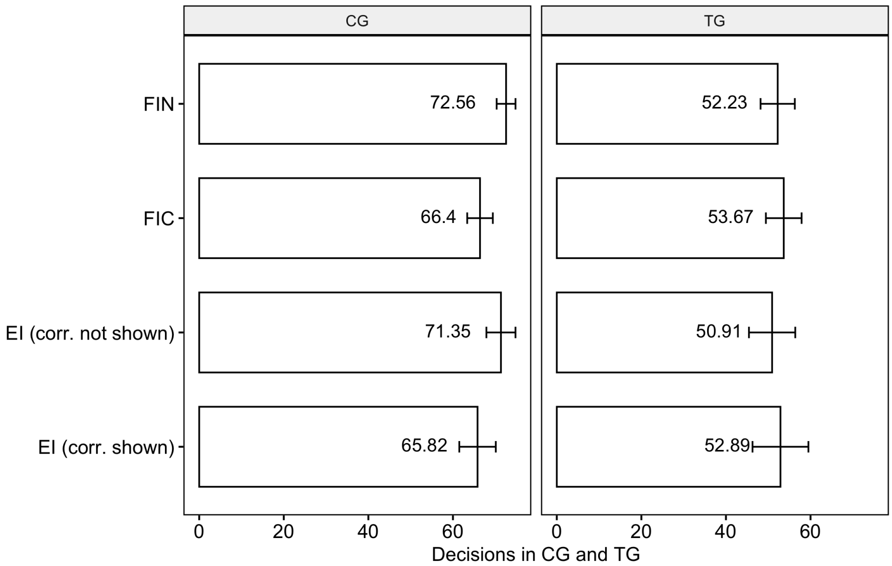

Figure A4: Overall decisions about Arkhangesk in *CG* and *TG*

Table A2: 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

A.3 Shapiro–Wilk normality tests of main DVs

Table A3: Pairwise Wilcoxon test for Arkhangelsk decisions

name	.y.	group1	group2	n1	n2	statistic	p	p
CG	value	EI (corr. shown)	EI (corr. not shown)	82	117	3905.0	0.025	0
CG	value	EI (corr. shown)	FIC	82	201	7886.5	0.569	1
CG	value	EI (corr. shown)	FIN	82	200	6561.0	0.008	0
CG	value	EI (corr. not shown)	FIC	117	201	13317.0	0.048	0
CG	value	EI (corr. not shown)	FIN	117	200	11700.0	1.000	1
CG	value	FIC	FIN	201	200	17269.5	0.014	0
TG	value	EI (corr. shown)	EI (corr. not shown)	82	117	5020.0	0.575	1
TG	value	EI (corr. shown)	FIC	82	201	8226.0	0.981	1
TG	value	EI (corr. shown)	FIN	82	200	8435.5	0.702	1
TG	value	EI (corr. not shown)	FIC	117	201	11214.0	0.486	1
TG	value	EI (corr. not shown)	FIN	117	200	11512.5	0.809	1
TG	value	FIC	FIN	201	200	20698.0	0.601	1

Table A4: Pairwise Wilcoxon test for Voronezh decisions only

name	.y.	group1	group2	n1	n2	statistic	p	p
CG	value	EI (corr. shown)	EI (corr. not shown)	82	117	5078.0	0.481	1
CG	value	EI (corr. shown)	FIC	82	201	8333.5	0.882	1
CG	value	EI (corr. shown)	FIN	82	200	8339.5	0.822	1
CG	value	EI (corr. not shown)	FIC	117	201	11198.5	0.476	1
CG	value	EI (corr. not shown)	FIN	117	200	11217.5	0.538	1
CG	value	FIC	FIN	201	200	20200.5	0.931	1
TG	value	EI (corr. shown)	EI (corr. not shown)	82	117	3682.0	0.005	0
TG	value	EI (corr. shown)	FIC	82	201	7304.0	0.129	0
TG	value	EI (corr. shown)	FIN	82	200	6732.0	0.017	0
TG	value	EI (corr. not shown)	FIC	117	201	13162.5	0.071	0
TG	value	EI (corr. not shown)	FIN	117	200	12123.5	0.586	0
TG	value	FIC	FIN	201	200	18682.5	0.215	0

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 A5: 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 A6: Pairwise two sample Wilcoxon tests for difference in beliefs

References

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. Publisher: SAGE Publications Sage UK: London, England.

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.

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. Publisher: Elsevier.

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

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.

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.

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.

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. Publisher: SAGE Publications Ltd.

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

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.

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

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.

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. Publisher: Elsevier.

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.

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. Publisher: SAGE Publications Inc.

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. Publisher: BioMed Central.

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. Publisher: Oxford University Press US.

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. _eprint: https://onlinelibrary.wiley.com/doi/pdf/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. Publisher: Public Library of Science.

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. Publisher: München: ifo Institut für Wirtschaftsforschung an der Universität München.

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. Publisher: Springer.

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

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.