City Research Online

City, University of London Institutional Repository

This is the published version of the paper.

This version of the publication may differ from the final published version.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

Identity and political corruption: a laboratory experiment

Maria Cubel[1] · **Anastasia Papadopoulou[2]** · **Santiago Sánchez-Pagés[3]**

Abstract
This paper explores the role of identity in voters' decision to retain corrupt politicians. We build up a model of electoral accountability with pure moral hazard and bring it to the lab. Politicians must decide whether to invest in a public project with uncertain returns or to keep the funds for themselves. Voters observe the outcome of the project but not the action of the politician; if the project is unsuccessful, they do not know whether it was because of bad luck or because the politician embezzled the funds. We run two treatments; a control and a treatment where subjects are assigned an identity using the minimal group paradigm. Our main result is that, upon observing a failed project, voters approve politicians of their same identity group significantly more often than in the control and compared to politicians of a different identity group. This is partially driven by a belief on same-identity politicians being more honest. We also observe that subjects acting as politicians embezzle funds less often than expected by the equilibrium prediction.

All authors have contributed equally to this work.

✉ Santiago Sánchez-Pagés
  santiago.sanchez-pages@kcl.ac.uk

  Maria Cubel
  maria.cubel@gmail.com

  Anastasia Papadopoulou
  anastasia.papadopoulou@bristol.ac.uk

1 Department of Economics, City University London, Northampton Square, London EC1V 0HB, UK

2 School of Economics, University of Bristol, The Priory Road Complex, Priory Road, Bristol BS8 1TU, UK

3 Department of Political Economy, King's College London, Bush House NE, London WC2B 4BG, UK

1 Introduction

One of the primary goals of elections is to hold politicians accountable for their past actions. Politicians should be less likely to engage in corrupt behaviour if they know they will be electorally punished for any wrong-doing (Ferejohn 1986). However, public sector corruption is still rife in democracies, as recent scandals demonstrate (e.g. UK parliamentary expenses scandal, Spain's EREs case). This is worrisome given the well-documented deleterious effect of corruption on economic growth, social welfare, and redistribution (e.g. Mauro 1995; Méon and Sekkat 2005; Gründler and Potrafke 2019).

A reason for the prevalence of corruption seems to be the frequent reluctance of citizens to vote out corrupt politicians.[1] This reduces the incentives of incumbents to refrain from using public office for private gain. Rundquist et al. (1977) argues that voters may not vote against corrupt politicians because they engage in an implicit trade-off between corruption and favourable policies. Fernández-Vázquez et al. (2016) provides empirical evidence in favour of this exchange argument between voters and politicians using data from the Spanish local elections. They also find evidence that the trading can be quite explicit, in the sense of voters expecting to receive direct side benefits from corruption.

An alternative explanation is in-group loyalty. Voters might be inclined to overlook corruption when the corrupt official shares their ethnic background or ideological beliefs. This inclination could stem from either a preference for politicians from their own group (taste-based) or an expectation that such politicians would be more likely to prioritize their interests. Using a survey experiment ran in Spain, Anduiza et al. (2013) find that voters tend to consider corruption scandals to be less serious when the politician involved belongs to the political party they favour. Another paper providing evidence on this partisan bias is by Eggers (2014), who documents a trade-off between voters' punishment of corrupt politicians' and their willingness to see their supported political party in office. Using data from the British expenses scandal, the author shows that electoral punishment of politicians involved in the scandal was weaker in constituencies where the electoral race was tighter. This phenomenon helps to explain why candidates in more ideologically polarized constituencies were more likely to be involved in the scandal in the first place.

The main goal of this paper is to study whether social identity affects the tolerance of voters towards corruption in the form of public funds embezzlement, i.e. the extraction of rents by a politician at the expense of voters (Persson et al. 2003; Ferraz and Finan 2011). We first build up a two-period model of electoral accountability with pure moral hazard. Politicians have the same level of competence but their actions when in office are unobservable to voters. The incumbent must decide whether to invest funds in a public project or to misappropriate them. The project has stochastic returns and the voter can only observe whether the project is successful or not. Hence, upon observing a failed project, the voter cannot know whether this was the result of bad luck or rent extraction. Then, the voter must decide whether to approve the incumbent or not.

---

[1] Evidence exists for Europe (Bågenholm 2013), Japan (Reed et al. 1996), and the United States (Welch and Hibbing 1997). For an overview, see Golden (2010).

We take this model to the lab to investigate the influence of identity on voters' decision to retain a potentially corrupt politician. We employed a between-subject design and conducted two treatments. In the control sessions, participants played in pairs and were randomly matched with strangers for 36 independent rounds; 18 not necessarily consecutive rounds in the role of citizen and the rest in the role of representative. In the identity treatment, participants were first assigned to a specific group before engaging in the main game. Because we employed the minimal group paradigm (Tajfel et al. 1971), the assigned group identities were constructed to be independent of personal characteristics that could potentially affect choices. In this treatment, participants played the same game as in the control treatment, with the only difference being that participants playing as citizens were informed about the identity of the representative they were matched with. Representatives did not know the identity of citizens and this was public knowledge. This eliminated any potential reciprocity effects based on identity, as citizens had no reason to exhibit reciprocal behaviour towards representatives of the same identity.

Our first result is that, after observing that the project is unsuccessful, citizens approve representatives around 7% less often when they do not belong to the same group compared to when they do. Our second result is that, conditional on the public project being unsuccessful, citizens approve representatives of their same identity group 11% more often compared to the control treatment. Our third result is that representatives extract rents less frequently than under the theoretical prediction. By investing the funds in the project, representatives earn around 28.3% less in expectation than the representatives who misappropriate the funds.

The observed differences in approval rates may be due to two reasons. One is pure pro-social preferences. As Chen and Li (2009) show, participants in dictator games tend to be more generous towards in-group members. In our experiment, in-group favouritism may operate in a similar manner since approving a representative awards them a payoff at no direct cost to the citizen. An alternative channel is differential beliefs about honesty.[2] If citizens expect same-identity politicians to be more honest, they might tend to reward that honesty by approving them more often. To explore these mechanisms we elicited beliefs from participants about the honesty of representatives upon having observed a failed project. We find that subjects believed that representatives of their same identity group were significantly less likely to have kept the funds compared to representatives of the other group.

A fruitful literature has employed experiments to study corruption and potential anti-corruption policies, focusing mainly on bribery. For example, Abbink et al. (2002) use a bribery game to study the effects of reciprocity, negative welfare effects, and severe punishment in corrupt behaviour. They find that reciprocation plays an important role; however, the introduction of severe penalties eliminates corruption. Barr and Serra (2009) show that externalities and framing effects affect bribing decisions, while Serra (2012) shows that combining top-down auditing and bottom-up monitoring can be an effective monitoring system to decrease corruption. In a repeated bribery

---

[2] In the rest of the paper we use the term "honest" to refer to representatives who do not misappropriate the funds. We follow previous papers in the political economy literature, such as Tirole (1996), Persson et al. (2003), Persson and Tabellini (2004), and Caselli and Morelli (2004), where the term "honest" is used to describe politicians who do not engage in embezzlement or rent extraction.

game using culturally diverse subjects, Banuri and Eckel (2015) show that punishment institutions against corruption need enforcement in order to have long lasting effects, while culture does not affect responsiveness to punishment in this framework. Also, Salmon and Serra (2017) study the effect of social enforcement in corruption and find that cultural differences affect social judgement. Di Zheng et al. (2021) study the effect of social ties on corruption in a bribery game. They find that the majority of subjects engage in bribing regardless of the presence of social ties (i.e., sharing the same identity with the subject receiving the bribe). Moreover, social ties lead to corrupt behaviour only when bribing is not allowed.

Part of this literature also studies corruption in the form of embezzlement. Barr et al. (2009) study service provider performance in an experiment ran with nursing students in Ethiopia and show that elected providers and better monitoring deter corrupt behaviour, while the effect of higher payments is not that strong. They also show that subjects' experience with the public sector affect their behaviour. Also, Banerjee et al. (2022) study embezzlement in a lab experiment ran in Kenya, showing that corruption is perceived as another cost of doing business without significant additional psychic costs. None of these papers on embezzlement study the influence of identity on this type of corruption.[3]

Our study is most closely related to two papers in the experimental political economy literature. On the one hand, Landa and Duell (2015) examine the effect of identity in politicians' effort choices and voters' behaviour in a setting with hidden competence and observable effort. In a lab experiment they find that voters show in-group favouritism by ignoring both politicians' effort and competence in in-group matches. Meanwhile, politicians exert some effort when they belong to the same group as voters, regardless of their competence and even if they expect to be re-elected. On the other hand, Solaz et al. (2019) study the effect of identity in corruption empirically and experimentally. In their empirical study, they take advantage of the publicity of a major corruption scandal in Spain to show how voters behaviour change after knowing that the incumbent party was corrupt. They find that voters punish the corrupt party, but partisanship eliminates this punishment. The authors also run a lab experiment inducing artificial identities to see if this in-group effect replicates in the lab. They show that in-group loyalty persists and that voters are more likely to select politicians of their same group even when it is known for certain that they were engaged in corruption.

In contrast with these two papers, we concentrate on how politicians are disciplined rather than selected. The two main goals of elections are to hold politicians accountable for past actions (sanctioning/retrospective voting) or to choose high-quality politicians (selection/prospective voting) (Ferejohn 1986; Fearon 1999; Besley 2005). Our focus is specifically on the sanctioning mechanism, which involves voters holding corrupt politicians accountable for their behaviour. While the two mechanisms can interact

---

[3] The role of identity in individual behaviour has been highlighted in the economic literature only recently. Akerlof and Kranton (2000) model the effect of identity ("a person's sense of self") in economic outcomes by incorporating it in agents' utility. The experimental evidence on the significant role of identity in individual decision is growing rapidly. As mentioned, Chen and Li (2009) provide evidence of the influence of identity on social preferences. Klor and Shayo (2010) highlight that group identity affects agents' preferences for redistribution biasing them in favour of their own group, while Cornaglia et al. (2019) show that individuals display increased competitiveness towards members of their same group.

with each other (Fearon 1999; Feltovich and Giovannoni 2015; Klingelhöfer 2021), experimental evidence suggests that sanctioning is significant for voters. Azfar and Nelson (2007) argue that voters hold corrupt politicians accountable, while Woon (2012) shows that voters decide retrospectively, even though they would have benefited from prospective voting. Moreover, Woon (2014) discusses that voters hold politicians accountable for their policy choices, even though politicians might ignore these incentives. Another difference is that we focus on moral hazard incentives by making politicians' actions unobservable and all of them equally competent. In addition, we remove any pivotality considerations and repeated interaction effects which could introduce potential confounds.[4]

The remainder of the paper is as follows. Section 2 contains the theoretical model we use to derive predictions for our experiment, while Sect. 3 describes the experimental design. Section 4 presents the results of the experiment and Sect. 5 concludes.

2 The model

2.1 Benchmark

This model adapts a two-period model of electoral accountability in Persson and Tabellini (2002) where public good provision has now uncertain returns.[5] There is one voter (she) with exogenous wealth y per period and one incumbent politician (he). There is an exogenous tax rate $\overline{\tau}$, so the tax revenue in each period is $\overline{\tau}y$. The incumbent must decide whether to pocket these revenues or to use them to fund a public project. Formally, the incumbent politician chooses in each period in office, $t = 1, 2$, the rent he extracts, $r_t \in \{0, \overline{\tau}y\}$. His payoff is $B_t + r_t$ where B_t are the rents from being in office in period t.

Given this choice, the level of public good provided at period t is given by $g_t(r_t, \theta) = \theta(\overline{\tau}y - r_t)$, where $\theta \in \{0, \overline{\theta}\}$ is a random variable which takes each value with probability $\frac{1}{2}$. The voter does not observe the rents extracted by the incumbent at the first period, r_1, but she observes the outcome of the project. If $g_1 = \overline{\theta}\overline{\tau}y$, the voter knows with certainty that the incumbent funded the project, i.e. $r_1 = 0$. If the voter observes $g_1 = 0$, she does not know whether the project failed due to randomness, i.e. $\theta = 0$, or because the incumbent kept the funds for himself, i.e. $r_1 = \overline{\tau}y$.

The voter's expected payoff at period t is thus

$$v_t = y(1 - \overline{\tau}) + \frac{g_t(r_t, \theta)}{2}, \tag{1}$$

where the first term is her private consumption and the second term is the expected outcome of the project.

The timing of the model is summarized in the following figure:

---

[4] Solaz et al. (2019) have groups of two candidates and three voters with the same composition through blocks of rounds whereas we have pairs of one representative and one voter re-matched in every round.

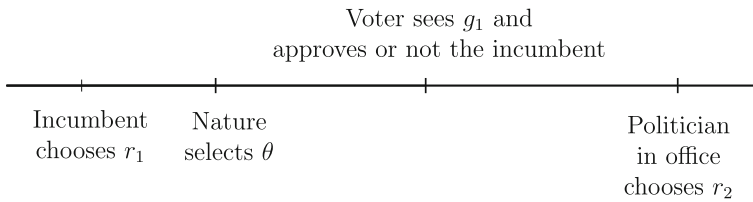[5] The authors refer to it as a model of career concerns.

Voter sees g_1 and
approves or not the incumbent

Incumbent	Nature		Politician
chooses r_1	selects θ		in office
			chooses r_2

Fig. 1 Timeline

After observing g_1, the voter must decide whether to approve the incumbent or not. If she approves him, then the incumbent reaches $t = 2$ and must choose the amount of rents to extract in that period, r_2. Since the politician has no incentive to refrain from extracting rents at that point, he chooses $r_2 = \overline{\tau} y$ so $g_2 = 0$. As a result, his payoff in the second period is $B_2 + \overline{\tau} y$, whereas the voter receives $v_2 = y(1 - \overline{\tau})$. If the voter does not approve the incumbent, he must step down and receives no payoff in the second period. The voter's payoff is $v_2 = y(1 - \overline{\tau})$ in that case too. This is equivalent to what the voter would receive if the incumbent were replaced by another politician who in turn would extract full rents.

Under these assumptions, the voter is indifferent between reelecting the incumbent or not. A multiplicity of equilibria emerges since the voter is indifferent between all her approval strategies. However, she is not indifferent among all possible equilibria. Because all approval rules are sequentially rational, punishments and rewards are credible and the voter can use them to discipline the incumbent. To be more specific, the voter's preferred outcome is that the incumbent chooses $r_1 = 0$. She can incentivize the incumbent to do this by using outcome-contingent approval strategies as follows.

We denote by a_H and a_L the probabilities with which the voter approves the politician when the project is successful, i.e. $g_1 = \overline{\theta} \overline{\tau} y$, and when the project fails, i.e. $g_1 = 0$, respectively. Given this strategy profile, the expected payoff for the incumbent politician at period 1 is

$$u_1 = B_1 + p\overline{\tau} y + \frac{1 - p}{2} a_H (B_2 + \overline{\tau} y) + \frac{1 + p}{2} a_L (B_2 + \overline{\tau} y), \qquad (2)$$

where p denotes the probability with which the incumbent embezzles the funds at $t = 1$. The second term in the above expression is the rents extracted in the first period, the third term is the expected value of re-election when the incumbent invests the funds and the project succeeds, and the last term is the expected value of re-election when $g_1 = 0$, either because the politician kept the funds or because he was honest but the project failed.

The incumbent chooses p to maximize (2). This optimization problem dictates that the incumbent refrains from rent extraction in period 1 if and only if

$$a_H - a_L \geq \frac{2\overline{\tau} y}{B_2 + \overline{\tau} y}. \qquad (3)$$

For the sake of exposition we assume that the voter approves the incumbent for sure when the project is delivered, i.e., $a_H = 1$; this is a sequentially rational choice for the voter following the arguments discussed above. In this case, the best response of the incumbent politician is to fund the project in the first period if and only if

$$a_L \leq \bar{a}_L \equiv 1 - \frac{2\bar{\tau}y}{B_2 + \bar{\tau}y}. \tag{4}$$

In other words, the incumbent can be disciplined if the approval probability when the project fails is low enough.

Proposition 1 *There exists a continuum of equilibria of the game where the politician is honest and the citizen approves him with probability $a_L \leq \bar{a}_L$ when she observes $g_1 = 0$ and with probability $a_H = 1$ otherwise. In addition, there exists a continuum of equilibria where the politician extracts rents, $a_H = 1$ and $a_L > \bar{a}_L$.*

Note that for the honest equilibrium to exist, $\bar{a}_L$ must be non-negative, that is, $B_2 \geq \bar{\tau}y$; in other words, the rents from being in office in the second period must be high enough compared to the rents the incumbent can extract in the first period.

2.2 Introducing identity

According to Tajfel and Turner (1986) the categorization of people in different groups can create in-group favouritism and out-group prejudice. In our context voters might condition their approval strategies on the identity of the politician when it is known to them.

To be more precise, let us rewrite the expected payoff of an incumbent politician belonging to group $i \in \{A, B\}$ at period $t = 1$. Now (2) becomes

$$u_{1i} = B_1 + p\bar{\tau}y + \frac{1-p}{2}a_H(B_2 + \bar{\tau}y) + \frac{1+p}{2}(\zeta_i a_L^s + (1 - \zeta_i)a_L^d)(B_2 + \bar{\tau}y), \tag{5}$$

where ζ_i is the probability with which the politician is matched with a voter of their same identity, and a_L^s and a_L^d are the voter's approval probabilities when she faces a same and a different identity politician respectively. For the sake of exposition, let us assume that voters approve politicians of either group with the same probability when the project succeeds, i.e. $a_H^d = a_H^s = a_H$. None of the results below hinge on this assumption.

Given (5), the best response of the incumbent is to be honest if and only if

$$\zeta_i a_L^s + (1 - \zeta_i)a_L^d \leq a_H - \frac{2\bar{\tau}y}{B_2 + \bar{\tau}y}. \tag{6}$$

Now, assume that voters feel in-group favouritism in the form of a "warm glow" when they do approve politicians of their same identity. Formally, voters' utility increases by a fixed amount $I > 0$ when they approve a politician of their same group regardless of the outcome of the project. In that case, the voter will always

approve the politician of their same group and punishments are no longer a credible threat, i.e. $a_L^s = a_H = 1$. Condition (6) thus boils down to

$$a_L^d \leq \overline{a}_L^d \equiv \frac{1}{1 - \zeta_i} \left(\frac{B_2 - \overline{\tau}y}{B_2 + \overline{\tau}y} - \zeta_i \right), \qquad (7)$$

so whether the incumbent is honest or not in equilibrium depends now on the proportion of voters in each identity group.

Proposition 2 *There exists a continuum of equilibria of the game with in-group favouritism where the politician is honest if and only if the citizen approves politicians of a different identity with probability $a_L^d \leq \overline{a}_L^d$ when she observes $g_1 = 0$ and approves with certainty otherwise. In addition, there exists a continuum of equilibria where the politician is dishonest, $a_H = a_L^s = 1$ and $a_L^d > \overline{a}_L^d$.*

For an equilibrium to exist in which incumbents of both identity groups are honest, $\overline{a}_L^d$ must be non-negative for both groups. Assuming a fully random matching so that ζ_i for $i \in \{A, B\}$ now denotes the proportion of voters in each group, an honest equilibrium exists if and only if

$$\frac{2\overline{\tau}y}{B_2 + \overline{\tau}y} \leq \zeta_i \leq \frac{B_2 - \overline{\tau}y}{B_2 + \overline{\tau}y} \quad \text{for} \quad i \in \{A, B\}. \qquad (8)$$

In words, this means that the proportion of voters in each identity group must not be too high or too small. Otherwise, incumbents in one of the groups would be too likely to meet a voter who will approve them unconditionally, eliminating any incentive to behave honestly. Note also that a necessary condition for an honest equilibrium to exist is again that the payoff from reelection should be high enough relative to the rents the incumbent can extract in the first period, i.e. $B_2 \geq 3\overline{\tau}y$.

2.3 From the theory to the lab

When implemented in the lab, the model described above may generate unwanted reciprocity effects. If a voter expects that reelecting the incumbent will make him more likely to be honest in the second period-even though that would run against his narrow self-interest- it would be rational for her to approve him when the project fails. This expectation of reciprocity may be reinforced by the presence of identity: an incumbent who sees himself approved despite a project failure may update up his beliefs about the voter having the same identity as his and increase his likelihood of being honest in $t = 2$ if in-group favouritism exists.

To rule out these effects, we implement in the lab a one period version of the game described above which produces the same theoretical predictions. In this version, the game ends after the voter's approval decision and the politician receives $B_2' = B_2 + \overline{\tau}y$ if approved. Assuming again that $a_H = 1$, the politician's expected payoff in the case without identity is

$$u_1 = B_1 + p\overline{\tau}y + \frac{1-p}{2}B_2' + \frac{1+p}{2}a_L B_2'. \qquad (9)$$

Now $\overline{a}_L \equiv 1 - \frac{2\overline{\tau}y}{B_2'}$ so for an honest equilibrium to exist $B_2' \geq 2\overline{\tau}y$.

If we were to introduce in-group favouritism as above, the one period version would require that for an honest equilibrium to exist voters must approve incumbents of their identity group with probability no greater than $\overline{a}_L^d \equiv 1 - \frac{1}{1-\varsigma_i}\frac{2\overline{\tau}y}{B_2'}$.

Consider the particular case where $B_2' = 2\overline{\tau}y$. This is the one we implement in our experiment. In that case, we can derive the following theoretical predictions:

Prediction 1: In absence of identity, an honest equilibrium exists if and only if $(a_L, a_H) = (0, 1)$.

Prediction 2: With in-group favouritism, an honest equilibrium does not exist, i.e. $\overline{a}_L^d < 0$.

3 Experimental design

Our experiment was pre-registered at "As Predicted".[6] Experimental sessions were ran at the Leicester Experimental Economic Laboratory (LExEcon) of the University of Leicester (UK) and at the Birmingham Experimental Economics Laboratory (BEEL) of the University of Birmingham (UK).[7] We ran nine sessions in total; six for the identity treatment and three for the control. The number of participants in each session was 14 and 16. Overall, we recruited 142 undergraduate and postgraduate students from several departments of the two Universities. The experiment was programmed in zTree (Fischbacher 2007).[8] The experiment ran in March and November 2019 and a session lasted approximately 40 min. The show up fee was £4 and the average payment across sessions was £14.9.

Subjects were randomly assigned to one of the two treatments, identity and control. The identity treatment had an initial additional stage where identity was induced using the Minimal Group Paradigm (MGP) as detailed below. Subjects played 36 independent rounds, half as "citizen" and half as "representative". We chose these words to provide some context and increase the external validity of the experiment. However, we deliberately refrained from using the word "politician" to avoid any strong connotations. We followed a similar approach when naming the actions in each role.

We followed the same strangers matching protocol as Woon (2012) with participants being re-matched in every round throughout the session whilst ensuring that all subjects had played in both roles by the twelfth round. At the beginning of each round each pair received 16 tokens, which served as tax revenue $\overline{\tau}y$. Subjects acting as representatives were told they were in charge of administering that amount, i.e decide whether to "keep" it or "send" it to the citizen. If the representative decided to keep the tokens, the citizen received 0 tokens. If the representative decided to send the tokens, the citizen received either double the amount (32 tokens) or nothing (0 tokens) with equal probability, i.e. $\theta = 2$. After observing the tokens received, but not the choice of the representative, the citizen decided whether to approve the representative or not. If the

---

[6] Registration #20716.

[7] All data collections were carried out after relevant ethics approval.

[8] See Table A.1 in Online Resource 1 for session details and Online Resource 3 for instructions.

citizen approved the representative, the latter received 32 tokens, i.e. $B_2' = 2\overline{\tau}y$. If the representative was not approved, he received nothing.[9] After this decision, the round ended. At the end of each round representatives were informed about the decision of the citizen and the total payoff the representative received in the round, while citizens were informed about their own payoff in the round. Our main interest is the comparison of approval decisions between in-group and out-group matches and the control.

In the identity treatment, citizens were informed about the identity of the representative they were matched with, while representatives received no information about the identity of the citizen. This allows us to pin down the effect of identity on citizens' decisions without worrying about possible reciprocity effects created by representatives knowing citizens' identity too. Our choice of letting citizens but not representatives know stems also from reality. It depicts better a scenario where voters know the identities of politicians who, in turn, know only imperfectly the ethnicity or ideology composition of their constituency.

Social identity was artificially induced and assigned at the beginning of each session of the identity treatment, following the widely-used minimal group paradigm. This minimal categorization is enough to create in-group favouritism and out-group animosity. The methodology has the advantage that artificially induced identities should be, in principle, orthogonal to subjects' characteristics which may in turn correlate with the behaviour under study. In other words it provides more control over the identity-formation process compared to natural identities (Charness and Chen 2020). However, according to Lane (2016), discrimination is higher when identity is induced artificially rather than naturally, i.e., it might be the case that the minimal group paradigm does not produce a lower bound of identity effects. The reasons for this can be various, including the fact that discrimination might be socially acceptable when the groups are not correlated with natural identities, like ethnicity or religion for example.[10] Acknowledging that both ways have their advantages and disadvantages, we believe that for the purposes of our experiment it was useful to abstract from any natural identities, in order to avoid as much as possible any correlation of the identity groups with specific political views or ideologies.

A prominent way to create minimal group identities in the lab is the use of the Klee-Kandinsky protocol (e.g. Tajfel et al. (1971), Chen and Li (2009)). Paul Klee and Wassily Kandinsky were two painters who worked closely roughly during the same time period in the same region in Germany and who remain relatively unknown to the general public till date. Art historians argue that their patterns are very similar.[11] As a result, it is to be expected that any categorization based on the taste for these painters should be as good as random.

---

[9] Notice that corruption is costly for citizens, as they receive 0 tokens, whereas they receive 16 tokens in expectation if the representative sends the money. Additionally, if citizens are more likely to approve the representative when the project is delivered, then the overall expected social welfare is higher in the absence of corruption.

[10] See Lane (2016) for a discussion on the possible explanations.

[11] "... similarities in scale, theme, motifs and even technique suggest a dialogue between the two artists. Their lives intersected at various times between 1911 and 1937, and their art responded to each other's." (Barnett 2015)

Specifically, our participants were shown five pairs of paintings. Each pair featured one painting by each painter. They were asked to choose the painting they liked most within each pair. They were then categorized as a "Klee" or a "Kandinsky" according to the majority of their choices. Following Chen and Li (2009) and Landa and Duell (2015), we also implemented an identity enhancement task.[12] Subjects were shown two additional paintings, one by Klee and one by Kandinsky. Members of each group had to identify the painting that belonged to the painter of their group. Subjects had to give two answers; their initial guess and their final answer after seeing the overall guess of the members of their identity groups. If the majority of members within a group identified the author correctly they all received 4 additional tokens.

In the model described in the previous section, we introduced in-group favouritism by assuming that identity entered directly in the utility of voters when approving a representative of their same group. However, in-group favouritism could also operate through voters' differential beliefs about the action of the representative upon observing that the project failed. According to the literature, voters may process information differently if they like the politicians involved in a corruption scandal (see for example Rahn (1993), Taber and Lodge (2006), and Anduiza et al. (2013)). In order to study whether in-group loyalty generates differential honesty beliefs, we elicited, at the end of the session, subjects' beliefs about the frequency of representatives keeping the funds. Subjects were paid 4 tokens if their guess was within a band of 5% points around the actual frequency.

At the end of each session, subjects completed a questionnaire on individual characteristics. It included questions on gender, age, field of study, religiosity, race/ethnicity, family income, ideology, volunteering activities, trust, and risk attitudes. In the identity sessions, subjects were also asked about their familiarity with the two painters and about the level of attachment with their group identity during the session.[13]

4 Results

4.1 Identity inducement

According to the minimal group paradigm, a categorization using artificially induced identities should be random and orthogonal to any personal characteristics. This also means that subjects should be split into two groups of roughly the same size. However, participants in our sample are divided into groups of quite unequal size. Subjects who preferred Kandinsky over Klee were 73.4% of the sample and this proportion is significantly different from 50% (proportion test, $p < 0.001$). Note however that this does not alter the theoretical predictions derived in section 2.3 as $B_2' = 2\overline{\tau}y$ implies that an honest equilibrium does not exist regardless of the proportion of subjects in each identity group ζ_i for $i = A, B$.

More worrisome perhaps is that we find evidence that identity correlates with some individual characteristics. The regression in Table 1 examines those individual charac-

---

[12] See Charness et al. (2007) for the importance of salience when using artificially induced identities.

[13] A descriptive analysis of the answers to the questionnaire can be found in Online Resource 1.

Table 1 Associations between personal characteristics and being assigned to the Kandinsky group

	LPM	Probit	Margins
Female	−0.1314	−0.6106*	−0.1406*
	(0.0663)	(0.3595)	(0.0796)
Age	−0.0630**	−0.2644**	−0.0609***
	(0.0163)	(0.1064)	(0.0232)
Centre ideology	−0.2471**	−1.1805***	−0.2794***
	(0.0711)	(0.4055)	(0.0787)
Right ideology	0.0002	0.1548	0.0249
	(0.1027)	(0.4507)	(0.0677)
Familiar with Klee	−0.3567	−1.7462	−0.4020
	(0.2161)	(1.2139)	(0.2976)
Familiar with Kandinsky	0.3136*	1.4539*	0.3347*
	(0.1541)	(0.7908)	(0.1975)
Key variable	Kandinsky	Kandinsky	Kandinsky
Degree FE	Yes	Yes	Yes
R-squared	0.35		
Pseudo R-squared		0.33	
Observations	94	86	86
Correct predictions	82%	83%	

Robust errors clustered by session. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

teristics that subjects reported in the questionnaire which were significantly associated with their choice of identity. Column (1) is a linear probability model on the likelihood of a subject being a "Kandinsky", while specifications (2) and (3) report the results of a probit model and the corresponding margins.

"Centre Ideology" and "Right Ideology" show the effect of a categorical variable that takes the value of 1 if the subject reported that their ideology is at the mid-point of a Likert ideology scale between 1 (extreme left) and 7 (extreme right) and the value of 2 if their ideology is between 5 and 7 in the same scale. "Familiar with Klee" and "Familiar with Kandinsky" take the value of 0 if the subject reported not being familiar with each painter, i.e. they chose 1 on a Likert scale between 1 (not familiar at all) and 7 (very familiar), and 1 otherwise.

Estimates from these regressions yield a significant negative association between being assigned to the Kandinsky group and being female, older, centrist, and unfamiliar with his work. The reasons behind these associations are of no interest to us but they imply that we must include these controls in all our main regressions. Otherwise, the correlation between personal characteristics and the "treatment" received might bias our estimates.
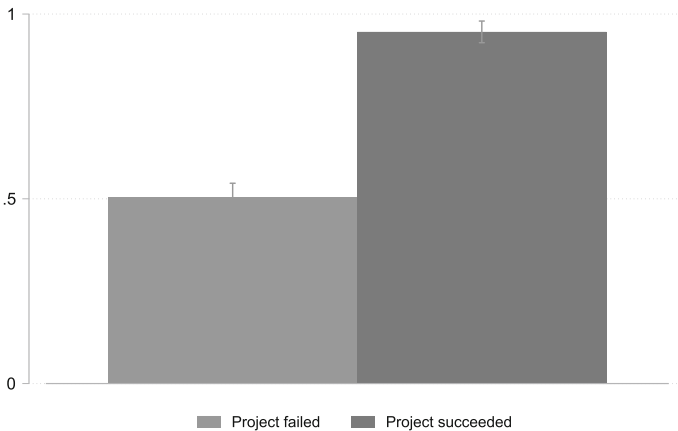
Fig. 2 Average approval by tokens received in the control sessions

4.2 Approval decisions

4.2.1 Control treatment

According to the theoretical predictions, if a_L, the approval probability when $g_1 = 0$, is low enough, representatives are honest and fund the project in equilibrium (Proposition 1). In particular, Prediction 1 states that under the values used in our experiment, an honest equilibrium exists if and only if citizens always approve the politician when the project succeeds and never if the project fails. For any other approval rates, the representative should pocket the funds.

Figure 2 shows that the observed approval rates in the control treatment do not match the values required to incentivize the representatives to behave honestly. Citizens approve far too often than they should after observing that they received no tokens, and less than they should after receiving the tokens. More specifically, the approval rate when citizens receive 0 tokens is 50.38%, whereas it is 95.17% when they receive 32 tokens. Even though 95.17% seems quite close to 100%, it is actually statistically different from full approval ($p < 0.001$).[14]

While these approval rates are not off-equilibrium, it is notable that citizens do not seem to aim to select an honest equilibrium. As discussed in Sect. 2, our game admits multiple equilibria because all retention rules are sequentially rational. However, by approving at such high rates after the project fails, citizens forego the opportunity to discipline representatives. This behaviour could be explained by a combination of preferences for social welfare and reciprocity concerns.[15]

---

[14] As we cannot run a proportion test if the proportion lies on the boundaries of the set (0, 1), we performed the test for 99.99%, used as an approximation.

[15] Social preferences such as social welfare concerns or inequity-aversion alone cannot explain the observed approval rates. Voters driven by social welfare should always approve, but only 13.4% of subjects did this. Similarly, inequity-averse voters should not approve when the project fails (to minimize unfavourable inequality) and always approve when it succeeds (to minimize favourable inequality). Thus, a more comprehensive approach is needed. In Online Resource 2, we apply Charness and Rabin (2002)'s social preferences
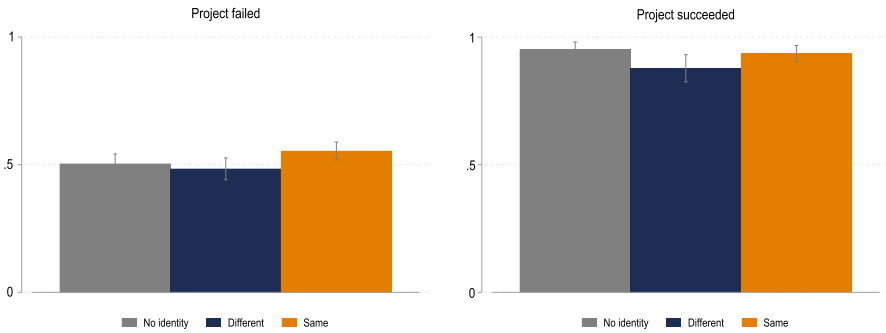
Fig. 3 Average approval by tokens received, treatment, and matchings

4.2.2 Identity treatment

Let us now discuss our main object of interest: approval rates in the identity treatment by type of matching. Figure 3 shows the approval rates by identity matches (control, different, and same) and by tokens received by the citizen. When citizens receive zero tokens, they approve more often representatives sharing their identity, and slightly less often the ones with a different identity.

Also, we observe that citizens tend to approve less often representatives with a different identity compared to the control even when the project is successful and they can be sure the representative was honest. However this difference is not statistically significant.[16]

Focusing on the difference on the approval decisions when the project failed, we run a linear probability model on the choice of the citizens controlling for the different treatments and matches.[17] Table 2 shows the result of the comparison between same identity matches and control ("No identity (control)"), and between same and different identity matches ("Different identity").[18] We use panel linear probability models with robust errors clustered by session to account for session heterogeneity.

More specifically, the first independent variable takes the value of 0 for the observations of the identity sessions and 1 otherwise. The second independent variable takes the value of 1 for the different identity matches and 0 otherwise. Moreover, specification (1) includes a control for round, a dummy for the location of the session (Birmingham or Leicester), and for whether the participant started the game by playing as a representative. Specification (2) adds the field of study as an additional control and all individual characteristics recorded in the questionnaire that correlate

---

Footnote 15 continued

model to the voter's approval decision, allowing for various forms of social preferences and reciprocity. This model accounts for high approval rates after project failures based on efficiency concerns or if the voter believes likely enough that the politician invested the funds.

[16] Tables A.3 and A.4 in Online Resource 1 provide the output of the relevant regressions.

[17] Random effects panel probit models are presented in Online Resource 1 for robustness (Tables A.5 and A.6).

[18] Table A.7 in Online Resource 1 replicates the same specification but using the comparison between control sessions and different identity matches.

Table 2 Effect of identity on approval choice after the project failed—panel linear probability models

	(1)	(2)
No identity (control)	−0.1109**	−0.1287**
	(0.0525)	(0.0503)
Different identity	−0.0674**	−0.0716**
	(0.0294)	(0.0291)
Round	−0.0021*	−0.0021*
	(0.0013)	(0.0013)
Birmingham	0.3240***	0.3651***
	(0.0329)	(0.0311)
Start as representative	0.0491	0.0635
	(0.0633)	(0.0738)
Centre ideology		0.0038
		(0.0592)
Right ideology		0.0017
		(0.1172)
Age		0.0274
		(0.0194)
Female		−0.0634
		(0.0473)
Key Variable	Approval (project failed)	Approval (project failed)
Degree FE	No	Yes
Observations	1971	1971

Robust errors clustered by session. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

with the choice of identity, except subjects' familiarity with the two painters, as these regressions include the control sessions.[19] We see that citizens in the control sessions approve representatives 11.1% less often compared to when they share identity with the representative in the identity treatment.

Moreover, we see that citizens who have different identity than the representatives approve them 6.7% less often compared to the same identity matches. We see also some learning effects, as citizens tend to approve less often the more rounds they play. Interestingly, the Birmingham dummy is highly statistically significant showing that subjects in the Birmingham sessions approved on average more often than the ones in Leicester. Even if we do not control for this effect, the differential coefficient between identity matches remains statistically significant at the 5% confidence level.

The findings so far can be summarized as follows:

Result 1 (Approval rates—Same vs. Different) *After receiving 0 tokens, citizens approve representatives more often when they share identity compared to when they do not.*

[19] Table A.8 in Online Resource 1 shows the same regression for the identity treatments only, controlling additionally for familiarity and identity. The differential coefficient between matches remains statistically significant and very similar in sign and magnitude.
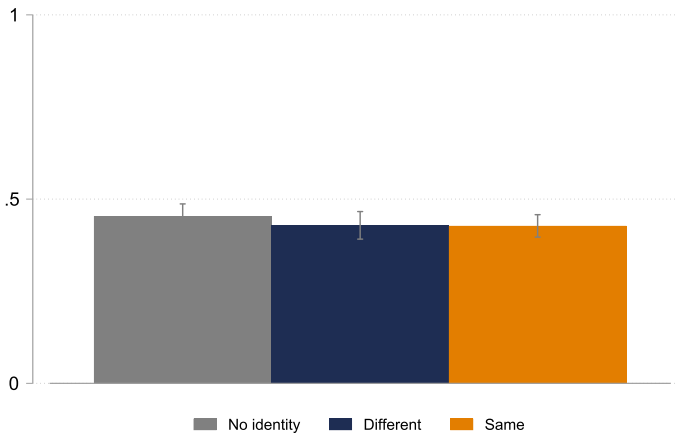
Fig. 4 Average sending by treatment and group matching

Result 2 (Approval rates—Same vs. No identity (control)) *After receiving 0 tokens, citizens approve representatives less often in the control compared to when they they share identity.*

4.3 Sending behaviour

4.3.1 Sending rates

Given citizens' high approval rates, purely self-interested representatives would be expected to keep the tokens. However, as shown in Fig. 4, they consistently exhibit significant sending rates across treatments and matches, indicating a clear preference for honesty. In the control treatment alone, representatives sent the tokens in 45.37% of the occasions. This behaviour has been observed in other accountability studies (e.g. Woon (2014)) and may be attributed to pro-sociality towards citizens (Meier 2007).

Result 3 (Sending rates—Honesty) *Given citizens' actual approval rates, representatives behave honestly more often than predicted.*

Let us now focus more on representatives' sending behaviour in the identity treatment. Prediction 2 states that, under the parameters we implement in the experiment, an honest equilibrium cannot exist if citizens show in-group favouritism and regardless of the identity composition of the set of potential voters. As a result, in the identity treatment we should observe lower sending rates than in the control. However, Fig. 4 shows that the sending rates across matches and treatments are almost identical. This suggests that representatives did not actually expect that in-group/out-group matches would change citizens' behaviour.

More specifically, in the identity treatment representatives keep the money on average in 57.21% of the occasions, whilst in 54.63% of the rounds in the control. These proportions are not significantly different at any conventional level for any comparison (control-different matchings, control-same matchings, same-different matchings)

Table 3 Effect of identity on sending decision of the representative—panel linear probability models

	(1)	(2)
No identity (control)	0.0371	0.0142
	(0.0423)	(0.0414)
Different identity	−0.0090	−0.0071
	(0.0116)	(0.0114)
Round	−0.0029**	−0.0029**
	(0.0013)	(0.0013)
Birmingham	−0.0808	−0.0708
	(0.0528)	(0.0789)
Start as representative	0.0710**	0.0794**
	(0.0315)	(0.0322)
Centre ideology		−0.0653
		(0.0712)
Right ideology		−0.1748***
		(0.0372)
Age		−0.0434***
		(0.0147)
Female		0.0758
		(0.0608)
Key variable	Sending	Sending
Degree FE	No	Yes
Observations	2556	2556

Robust errors clustered by session. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

with the p-values of the two-sample proportion tests being higher than 0.1 in all cases. Moreover, recall that in the identity treatment citizens know the identity of the representatives they are matched with, but not vice versa. As a result, we expected no difference between the different and same matches. This provides evidence of the internal validity of our experiment.

Table 3 below presents the regression results showing that there is indeed no statistically significant difference neither in the average sending rates across treatments nor between same and different matches.

The first specification in Table 3 includes the main independent variables controlling for the round, the sessions ran at the experimental lab in Birmingham, and the effect of playing in the first 6 rounds as a representative.[20] In specification (2) we add those individual characteristics that played a role in the identity choice.[21] We see that in both cases there is no difference in the choice of representatives among treatments. To be more specific, the coefficient of the control compared to the same identity matches is positive in both specifications, but it is far from significant. The coefficient

---

[20] Table A.9 in Online Resource 1 includes the corresponding panel probit models for robustness.

[21] In Table A.10 of Online Resource 1 we include the same regression but only for the identity sessions, controlling additionally for familiarity with the painters and identity.

Table 4 Representatives' choices by citizens' decisions

Rep's choice	Citizen's choice		
	Not approved (0)	Approved (32)	Total
Keep (16)	691	749	1440
Send (0)	301	815	1116
Total	992	1564	2556

of the differential between the different and the same matches is again not statistically significant. There seems to exist some learning, as the more periods subjects play the more often they keep the money, getting closer to the theoretical prediction. Also, we see that subjects who played as representatives in the first 6 rounds tend to send more money compared to those starting as citizens. This may be due to the fact that, once subjects play as citizens, they update their beliefs about the approval rates and keep the tokens more often when they later play as representatives. Subjects' political ideology and age play also a role in their decision when they play as representatives. Older participants and participants who self-report to be conservative are more likely to keep the tokens. However, including these individual characteristics does not affect our main result about representatives' behaviour:

Result 4 (Sending rates) *There is no difference in representatives' sending behaviour across treatments and matches.*

4.3.2 Representatives' earnings

In both treatments, conditional on the actual approval rates of the citizens, representatives forgo significant earnings by sending the tokens. That would suggest that representatives suffer a cost from not keeping the tokens for themselves. In order to measure this cost, we next calculate the loss in representatives earnings due to honesty by using the observed approval frequencies.

Table 4 presents the frequencies of the four possible combinations of representatives' and citizens' choices (keep/send by approve/not approve) including all sessions (recall representatives were not aware of the citizen's identity in the identity treatment). The numbers in parentheses are the amount of tokens a representative earns with that choice; their combination yields his final earning. For instance, if a representative kept the funds and the citizen approved him, the representative earned $16 + 32 = 48$ tokens. Using this information, we can calculate the expected earnings of a representative who keeps or sends the money. If a representative kept the money (16 tokens), he received on average $(1,440 \times 16 + 749 \times 32)/(1,440) = 32.64$ tokens, whereas he received $(815 \times 32)/(1,116) = 23.4$ tokens in expectation if he sent the funds. In other words, an honest representative forgoes 9.24 tokens or £4.62, which represents 28.3% less than the expected earning of a representative who kept the funds.

Table 5 Honesty beliefs vs actual conditional honesty rates when project fails

	Average honesty beliefs		Average conditional honesty		Beliefs ≠ Honesty
	Obs.	%	Obs.	%	(*p*-values)
Same	47	41.92	787	25.92	0.0000
Different	47	34.28	527	26.95	0.0671
Control	48	38.56	657	28.16	0.0055

4.4 Beliefs

One remaining question is why citizens, upon observing a failed project, approve politicians of their same identity significantly more often than predicted. There are two potential reasons for this. The first one is that participants' in group favouritism operates through pure pro-social preferences, independent of the representative's action (Chen and Li 2009). If approving the representative awards the citizen a positive payoff, and participants include in their utility function the welfare of their fellow group members, they will be more inclined to approve in-group representatives. On the other hand, it may just be that participants hold more optimistic beliefs about in-group representatives and they think that they are more likely to be honest, or in other words, they are more willing to give them the benefit of the doubt. In order to check whether the second channel might be relevant, we elicited beliefs about honesty rates by asking participants the following question at the end of the experiment:

> "From all the rounds in which you played as a citizen, (the representative was a Klee/Kandinsky), and you received 0 tokens, what is the percentage of occasions in which you believe the representative kept the tokens for themself?"

This question was incentivized. Participants earned 4 additional tokens if their answers were in a 5 percentage points band around the true percentage of representatives who kept the funds.

In the control treatment, participants were asked about all rounds. In the identity treatment, half of the subjects in each identity group answered the question for representatives of their same identity and the other half for out-group representatives. With this we forwent some data but we avoided any experimenter demand effect that could have arisen from asking about the honesty of both types of representatives.

Recall that representatives did not know the citizen's identity and this was public knowledge. Hence, this question should elicit pure priors about the behaviour of same and different identity participants.

Table 5 below presents descriptive statistics for these elicited beliefs, providing the complement of the average value of subjects' answer to the question on beliefs in the control treatment and in the identity treatment for same and different matchings. It also includes the average actual proportion of the representatives honesty decisions by treatment and matches when citizens receive 0 tokens.

The first observation from this table is that participants believed that representatives were significantly more honest than they actually were on average across treatments

and matches. In all three cases, the difference between the average honesty belief and the actual average conditional honesty of representatives across treatments and matches is statistically significant. This helps to explain the high approval rates we observe when citizens received zero tokens compared to the equilibrium prediction.

Participants overestimated the honesty of representatives to the largest extent in the same identity matches. Actual average honesty rates were very similar across matches but citizens trusted representatives of their same identity significantly more compared to representatives with different identity. In order to see if this effect is significant we run a number of regressions and present the results in Table 6. In all three specifications the dependent variable is the complement of the elicited beliefs measured in percentage. The coefficient for the difference between average honesty beliefs for same and different identity becomes statistically significant after controlling for subject's honesty as representative. The relationship between individual honesty and subjects' belief on the honesty of representatives is highly significant and positive. That is, the more honest a subject was when playing as representative the more honest they believed other representatives were.[22] In columns (2) and (3) we also control for the percentage of occasions in which the participant approved the representative after receiving zero tokens ("Average Conditional Approval"). There is no evidence that participants who approved more often had different beliefs. Moreover, column (3) includes controls for participants' characteristics. We see a correlation only with subjects' ideology; subjects who self-reported being centrists tended to believe that representatives were more honest. These results are largely robust to alternative specifications.[23]

To sum up, elicited beliefs suggest that the differences we observe in approval rates could be at least partially due to different priors about honesty. Upon observing a failed project, participants believe that the representatives of their same identity are more honest than the representatives from the other identity group. Citizens might have wanted to reward representatives they believed were honest by approving them. Because they expected representatives of their same identity to be more honest, citizens might have approved them more often. To verify this, we analysed the effect of beliefs on approval decisions; however, we do not detect any statistically significant relationship in our data.[24]

The result below summarizes the analysis.

---

[22] There is evidence that people's beliefs about others can be affected by their own behaviour. See for example Erkal et al. (2022).

[23] In Tables A.11 and A.12 of Online Resource 1 we include the same regressions for the identity sessions, controlling additionally for familiarity with the painters and identity. In Table A.13 we include also a regression for all sessions controlling for experience of receiving 0 tokens as voters.

[24] One of the specifications in Table A.14 of Online Resource 1 limits the sample to observations with same (different) matchings where the subject replied to the belief question for same (different) identity representatives. No significant effect is observed there, although this analysis is constrained by the design of our belief elicitation question. This finding is not unique to our study. For instance, Benistant and Villeval (2019) also observed differences in subjects' beliefs depending on others' group identity, but no corresponding impact on behaviour. In their research, no discernible effect of identity on subjects' behaviour is found. In contrast, our study reveals in-group favouritism in subjects' approval rates, although this association lacks statistical significance.

Table 6 Effect of identity on elicited honesty beliefs—Linear probability models

	(1)	(2)	(3)
No identity (control)	−2.1208	−4.8218*	−5.2295
	(3.4860)	(2.5174)	(3.4406)
Different identity	−6.5260	−8.3866**	−8.8149**
	(4.1697)	(3.6002)	(3.5811)
Birmingham	−3.6627	0.5446	1.2960
	(4.7279)	(3.2199)	(3.9175)
Start as representative	5.2172*	2.5111	3.3303
	(2.6566)	(2.0599)	(2.8479)
Average Honesty (%)		0.3593***	0.3687***
		(0.0519)	(0.0477)
Average conditional approval (%)		−0.0280	−0.0281
		(0.0589)	(0.0678)
Centre ideology			8.0717**
			(3.3551)
Right ideology			3.2747
			(3.7559)
Age			−0.4417
			(0.9210)
Female			0.8503
			(5.3605)
Key variable	Honesty beliefs (%)	Honesty b eliefs (%)	Honesty beliefs (%)
Degree FE	No	No	Yes
Observations	142	142	142

Robust errors clustered by session. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Result 5 (Beliefs) *Citizens who shared identity with the representative believed that it was less likely that the representative kept the money when they received 0 tokens compared to when they did not share identity with the representative.*

5 Conclusion

There is substantial evidence showing that voters are often reluctant to remove corrupt politicians from office. The evidence also suggests that in-group favouritism is one important driver of this phenomenon. In this paper, we brought to the lab a pure moral-hazard model of electoral accountability with stochastic public good provision. We found that social identity is an important factor in voters' reluctance to vote out possibly corrupt politicians. Specifically, voters who observe an unsuccessful project tend to approve more often politicians who belong to their same identity group than those who do not or when identity is absent.

This result is in line with the empirical evidence that partisanship plays an important role in corruption. We corroborate that in-group favouritism operates independently of ideological or reciprocity (quid pro quo) considerations. We have shown that, even when identity is artificially induced, citizens tend to believe politicians of their same identity are relatively more honest and retain them in office more often. We leave for future research to explore the influence of identity on the selection of candidates with different abilities.

Data availibility The data that supports the findings of this study are available from the corresponding author upon request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Abbink, K., Irlenbusch, B., Renner, E.: An experimental bribery game. J. Law Econ. Org. **18**, 428–454 (2002)

Akerlof, G.A., Kranton, R.E.: Economics and identity. Q. J. Econ. **115**, 715–753 (2000)

Anduiza, E., Gallego, A., Muñoz, J.: Turning a blind eye: experimental evidence of partisan bias in attitudes toward corruption. Comp. Pol. Stud. **46**, 1664–1692 (2013)

Azfar, O., Nelson, W.R.: Transparency, wages, and the separation of powers: an experimental analysis of corruption. Public Choice **130**, 471–493 (2007)

Bågenholm, A.: Throwing the rascals out? The electoral effects of corruption allegations and corruption scandals in Europe 1981–2011. Crime Law Soc. Chang. **60**, 595–609 (2013)

Banerjee, R., Boly, A., Gillanders, R.: Is corruption distasteful or just another cost of doing business? Public Choice **190**, 33–51 (2022)

Banuri, S., Eckel, C.: Cracking down on bribery. Soc. Choice Welfare **45**, 579–600 (2015)

Barnett, V.E.: To my dear friend of many Years-Klee and Kandinsky's works on paper, 1911–1937, in Klee & Kandinsky: neighbors, friends, rivals. Prestel Verlag, Munich, Germany (2015)

Barr, A., Lindelow, M., Serneels, P.: Corruption in public service delivery: an experimental analysis. J. Econ. Behav. Org. **72**, 225–239 (2009)

Barr, A., Serra, D.: The effects of externalities and framing on bribery in a petty corruption experiment. Exp. Econ. **12**, 488–503 (2009)

Benistant, J., Villeval, M.C.: Unethical behavior and group identity in contests. J. Econ. Psychol. **72**, 128–155 (2019)

Besley, T.: Political selection. J. Econ. Perspect. **19**, 43–60 (2005)

Caselli, F., Morelli, M.: Bad politicians. J. Public Econ. **88**, 759–782 (2004)

Charness, G., Chen, Y.: Social identity, group behavior, and teams. Ann. Rev. Econ. **12**, 691–713 (2020)

Charness, G., Rabin, M.: Understanding social preferences with simple tests. Q. J. Econ. **117**, 817–869 (2002)

Charness, G., Rigotti, L., Rustichini, A.: Individual behavior and group membership. Am. Econ. Rev. **97**, 1340–1352 (2007)

Chen, Y., Li, S.X.: Group identity and social preferences. Am. Econ. Rev. **99**, 431–57 (2009)

Cornaglia, F., Drouvelis, M., Masella, P.: Competition and the role of group identity. J. Econ. Behav. Org. **162**, 136–145 (2019)

Di Zheng, J., Schram, A., Doğan, G.: Friend or foe? Social ties in bribery and corruption. Exp. Econ. **24**, 854–882 (2021)

Eggers, A.C.: Partisanship and electoral accountability: evidence from the UK expenses scandal. Q. J. Polit. Sci. **9**, 441–472 (2014)

Erkal, N., Gangadharan, L., Koh, B.H.: By chance or by choice? Biased attribution of others' outcomes when social preferences matter. Exp. Econ. **25**, 413–443 (2022)

Fearon, J.D.: Electoral accountability and the control of politicians: selecting good types versus sanctioning poor performance. Democr. Account. Represent. 55–97 (1999)

Feltovich, N., Giovannoni, F.: Selection vs. accountability: an experimental investigation of campaign promises in a moral-hazard environment. J. Public Econ. **126**, 39–51 (2015)

Ferejohn, J.: Incumbent performance and electoral control. Public Choice **50**, 5–25 (1986)

Fernández-Vázquez, P., Barberá, P., Rivero, G.: Rooting out corruption or rooting for corruption? The heterogeneous electoral consequences of scandals. Polit. Sci. Res. Methods **4**, 379–397 (2016)

Ferraz, C., Finan, F.: Electoral accountability and corruption: evidence from the audits of local governments. Am. Econ. Rev. **101**, 1274–1311 (2011)

Fischbacher, U.: z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. **10**, 171–178 (2007)

Golden, M.A.: Some puzzles of political corruption in modern advanced democracies, in in democracy and accountability: globalized political responsibility, pp. 184–199. Fukosha, Toyko (2010)

Gründler, K., Potrafke, N.: Corruption and economic growth: new empirical evidence. Eur. J. Polit. Econ. **60**, 101810 (2019)

Klingelhöfer, J.: Competitive elections, incumbency advantage, and accountability. Econ. Theor. **71**, 1397–1428 (2021). https://doi.org/10.1007/s00199-021-01370-9

Klor, E.F., Shayo, M.: Social identity and preferences over redistribution. J. Public Econ. **94**, 269–278 (2010)

Landa, D., Duell, D.: Social identity and electoral accountability. Am. J. Polit. Sci. **59**, 671–689 (2015)

Lane, T.: Discrimination in the laboratory: a meta-analysis of economics experiments. Eur. Econ. Rev. **90**, 375–402 (2016)

Mauro, P.: Corruption and Growth. Q. J. Econ. **110**, 681–712 (1995)

Meier, S.: A survey of economic theories and field evidence on pro-social behavior. In: Frey, B.S., Stutzer, A. (eds.) Economics and psychology: a promising new cross disciplinary field. The MIT Press, Cambridge (2007)

Méon, P.-G., Sekkat, K.: Does corruption grease or sand the wheels of growth? Public Choice **122**, 69–97 (2005)

Persson, T., Tabellini, G.: Constitutions and economic policy. J. Econ. Perspect. **18**, 75–98 (2004)

Persson, T., Tabellini, G., Trebbi, F.: Electoral Rules and Corruption. J. Eur. Econ. Assoc. **1**, 958–989 (2003)

Persson, T., Tabellini, G.E.: Political economics: explaining economic policy. MIT press, Cambridge (2002)

Rahn, W.M.: The role of partisan stereotypes in information processing about political candidates. Am. J. Politi. Sci. **37**, 472–496 (1993)

Reed, S.R., et al.: Political corruption in Japan. Int. Soc. Sci. J. **48**, 395–405 (1996)

Rundquist, B.S., Strom, G.S., Peters, J.G.: Corrupt politicians and their electoral support: some experimental observations. Am. Politi. Sci. Rev. **71**, 954–963 (1977)

Salmon, T.C., Serra, D.: Corruption, social judgment and culture: an experiment. J. Econ. Behav. Org. **142**, 64–78 (2017)

Serra, D.: Combining top-down and bottom-up accountability: evidence from a bribery experiment. J. Law Econ. Org. **28**, 569–587 (2012)

Solaz, H., De Vries, C.E., de Geus, R.A.: In-group loyalty and the punishment of corruption. Comp. Pol. Stud. **52**, 896–926 (2019)

Taber, C.S., Lodge, M.: Motivated skepticism in the evaluation of political beliefs. Am. J. Polit. Sci. **50**, 755–769 (2006)

Tajfel, H., Billig, M.G., Bundy, R.P., Flament, C.: Social categorization and intergroup behaviour. Eur. J. Soc. Psychol. **1**, 149–178 (1971)

Tajfel, H., Turner, J.: The social identity theory of intergroup behavior. In: Worchel, S., Austin, W. (eds.) The psychology of intergroup relations, pp. 7–24. Nelson Hall, Chicago (1986)

Tirole, J.: A theory of collective reputations (with applications to the persistence of corruption and to firm quality). Rev. Econ. Stud. **63**, 1–22 (1996)

Welch, S., Hibbing, J.R.: The effects of charges of corruption on voting behavior in congressional elections, 1982–1990. J. Politi. **59**, 226–239 (1997)

Woon, J.: Democratic accountability and retrospective voting: a laboratory experiment. Am. J. Politi. Sci. **56**, 913–930 (2012)

Woon, J.: An experimental study of electoral incentives and institutional choice. J. Exp. Pol. Sci. **1**, 181–200 (2014)