TI 2025-013/I
Tinbergen Institute Discussion Paper

# The Role of Opinion Polls in Coordination Amongst Protest Voters: An Experimental Study

*Oliver Feltham*[1]
*Arthur Schram*[2]
*Randolph Sloof*[3]

1 University of Amsterdam, Tinbergen Institute
2 University of Amsterdam, Tinbergen Institute
3 University of Amsterdam, Tinbergen Institute

# The Role of Opinion Polls in Coordination Amongst Protest Voters: An Experimental Study*

Oliver Feltham[†], Arthur Schram[‡], Randolph Sloof[§]

February 2025

**Abstract**

In an election, protest voters signal their discontent with the party they traditionally support in different ways. This paper examines a specific form of protest voting in which voters choose an anti-mainstream party over their true first preference, the mainstream party, as a way to signal discontent with mainstream policies or influence future policy decisions. Protest voters face a trade-off stemming from a coordination problem. Too few protest votes mean that the strength of the protest is insufficient to affect the mainstream's policies; too many protest votes may result in an anti-mainstream victory, which is a sub-optimal outcome for the protest voter. One way to address this coordination problem is through opinion polls. In this context, polls serve a dual purpose: they provide information about the challenges protest voters face (information channel) and function as a coordination mechanism, allowing voters to adjust their behaviour based on poll results to resolve the coordination problem (coordination channel). We test, experimentally, the extent to which each of these channels increases the likelihood that the protest is successful and find that both channels are significant.

*We are grateful to Giovanni Andreottola, Katharina Brütt, Gary Charness, Patrick Francois, Ben Greiner, Jan Hausfeld, Alejandro Hirmas, Konstantinos Ioannidis, Natalie Lee, Andrea Mattozzi, Theo Offerman, Dana Sisak, Joep Sonnemans, Robert Sugden, Junze Sun, Jean-Robert Tyran, Linh Vu and Andreas Ziegler for comments and discussions. We also thank seminar participants at the IOEA Spring School 2022, CCC meetings 2022 and 2024, MPI-Bonn meeting 2023, ESA World Meeting 2023, IMEBESS conference 2023 and seminar participants at the universities of Amsterdam (2022), Torino (2024) and Catania (2024). The Economics & Business Ethics Committee (University of Amsterdam) approved this project on 27 June 2022 with reference number EC 20220627010610.

[†]CREED, Amsterdam School of Economics and Tinbergen Institute. O.J.Feltham@uva.nl.

[‡]CREED, Amsterdam School of Economics and Tinbergen Institute, A.J.H.C.Schram@uva.nl.

[§]Amsterdam School of Economics and Tinbergen Institute. R.Sloof@uva.nl.

# 1 Introduction and Motivation

Understanding voters' choices is fundamentally important to understanding what drives election outcomes. Economic models of voter behaviour typically assume that voters are rational and vote for the party whose expected government action provides them with the highest utility for the subsequent period of office (Downs, 1957). Maintaining this assumption on voters' motivations is, however, too limiting to understand the full range of observed voting behaviour (Piketty, 2000; Castanheira, 2003). For example, it is difficult to argue that a voter who spoils their ballot in an election does so because they think this will result in a government that provides them with a higher utility than casting a vote for one of the available parties or not going to the polling station in the first place. Instead, such a voter may intend to register discontent towards factors such as the current economic circumstances (Bowler & Lanoue, 1992); the candidates running for public office (Adams et al., 2006); or some aspect of the policy programme being offered by the incumbent government (Franklin et al., 1994; Myatt, 2017).

It is common in the literature to refer to voters who use their vote to register their discontent in some way as protest voters. Actions taken by protest voters at the ballot box include spoiling a ballot, casting a vote for 'none of the above' or voting for a less preferred party in order to send their most preferred party a signal of dissatisfaction (Kselman & Niou, 2011; Alvarez et al., 2018). In this paper, we limit our attention to this latter form of protest voting, dubbed tactical protest voting in Alvarez et al. (2018), but drop the 'tactical' prefix when referring to protest voting, voters and votes.

Such protest voters can be understood as having dual objectives: (i) they want their most preferred candidate (the mainstream candidate[1]) to actually win the election and (ii) they want the less preferred candidate (the anti-mainstream candidate) to obtain a number of votes that exceeds a critical threshold (Myatt, 2017). The underlying logic behind the second component of this incentive structure is that the mainstream candidate responds to the protest voters' signal of dissatisfaction only if the strength of this feeling in the electorate as a whole is sufficiently high (Myatt, 2017). Where the number of votes for the anti-mainstream exceeds the critical threshold, the mainstream candidate responds by changing their policy offering: this could involve dropping the current policy programme, or adopting a new policy, perhaps one that is being propagated by the anti-mainstream candidate (Myatt, 2017).

Protest voters do not, therefore, want the anti-mainstream candidate to obtain so many votes that they actually win the election. If protest votes help the anti-mainstream candidate win the election,[2] the protest fails because a mainstream victory remained the protest voter's

[1]Depending on the context, this could be the governing party, but we also permit an interpretation in which the mainstream candidate represents the constitutional, legal or political status quo. For example, we argue that our setup applies to referenda where the mainstream candidate is interpreted as being the current status of affairs, or the option that is being propagated by the political establishment.

[2]In other words, if the genuine underlying level of support for the anti-mainstream was not sufficiently high to

genuinely most preferred option and we refer to such an event as an electoral accident (Louis et al., 2022). We refer to the region within which the protest succeeds as the *critical region*. Depending on the configuration of preferences in the electorate, the set of protest voters may then face an anti-coordination[3] problem (Myatt, 2017) in which the optimal outcome is for a strict subset to vote for the anti-mainstream candidate.

There have been various examples in recent years of voters engaging in behaviour consistent with this kind of protest voting. One form that this can take is voting for parties outside of the established choices. A prominent example is given by votes for the United Kingdom Independence Party (UKIP) or the Brexit Party being cast by traditional Conservative Party voters to signal discontentment with the UK's membership of the European Union (Alvarez et al., 2018). Protest votes cast for UKIP can be conceived of as having been successful in that those votes resulted in the Conservatives carrying out a policy concession in their agreement to a referendum on EU membership (Myatt, 2017; Schimpf, 2021). Subsequent election results have resulted in a collapse of support for UKIP, with many switching back to the Conservatives (Heath & Goodwin, 2017); this is consistent with the explanation that some voters have engaged in protest voting.

Alternatively, one might also consider events such as the 2014 Scottish independence referendum (Myatt, 2017) and the 2016 UK EU membership referendum (Louis et al., 2020) as instances under which some voters engaged in protest voting. In the Scotland example, some voters may have cast their vote for 'Yes'[4] to encourage the UK Government to increase the level of devolved powers provided to the Scottish Parliament (Myatt, 2017). In the EU referendum example, Fetzer (2019) indicates that the UK Government's policy of austerity that followed the financial crisis encouraged higher support for the 'Leave'[5] campaign, causing some voters to vote to leave the EU to express dissatisfaction with the status quo (Alabrese & Fetzer, 2018). Fetzer (2019) argues that many voters who supported 'Leave' did so because the policy of austerity encouraged the development of anti-establishment preferences and that this policy alone may have been sufficient to effect the outcome of the referendum. If these arguments hold true, an electoral accident happened in the case of Brexit, but not for Scottish independence. These examples highlight how protest voting can affect the political landscape; therefore, expanding our knowledge on the conditions under which protest voting is more prevalent contributes to

enable the anti-mainstream to win without protest voter support.

[3]An anti-coordination game, for example the game of chicken or the snowdrift game introduced by Sugden (1986), is one in which a player's best response is to play an action that differs to that of at least one of their partners'. This contrasts with a coordination game, in which players' mutual best responses are to play the same action. For our purposes, we use the terms anti-coordination and coordination interchangeably and refer to coordination more broadly, as the notion that voters effectively cast sufficient votes for the anti-mainstream candidate to ensure that the protest is successful.

[4]In the 2014 Scottish independence referendum, the question asked was whether Scotland should become an independent country, with a vote for 'Yes' indicating a vote in favour of the proposition.

[5]In the 2016 EU referendum, voters could vote to 'Remain' in or 'Leave' the EU.

our understanding of electoral outcomes.

One variable that will likely impact a voter's decision to engage in protest voting is the level of information that she holds about the distribution of preferences and voting intentions in the electorate in which she participates. Pre-election opinion polls provide such information and have been shown to act as a coordination device in scenarios in which voters have incentives to vote strategically (Forsythe et al., 1993, 1996; Tyszler & Schram, 2016). In such situations, opinion polls can provide two functions: to reveal information on the underlying distribution of preferences in the electorate and to provide a device that enables voters to coordinate their voting decisions to try to effect a more preferable outcome (Andonie & Kuzmics, 2012). The aforementioned studies consider cases where coordination is needed to defeat a Condorcet loser. Instead, we study a scenario where voters face a different (anti-)coordination problem, namely, ensuring that the number of protest votes falls within the critical region.

Identifying the effect of a specific institution on the incidence of successful protest voting in actual elections can be a difficult task; empirical studies based on observational field data will typically be unable to isolate the impact of any single specific institution. To overcome this shortcoming, we employ an experimental approach to test how opinion polls impact the success of protest voting. Our paper is, perhaps, most closely related to Louis et al. (2022), which is the first experimental contribution that tests the theory of protest voting developed by Myatt (2017). Whereas Louis et al. (2022) focus on how protest voting is affected by the popularity of a protest and the protest's salience, our paper addresses a different question: that of the effect of opinion polls on protest voting.

We study a stylised scenario in which voters are assigned one of three types: mainstream core supporters, anti-mainstream core supporters and protest voters. Core supporters are defined as those who definitely vote for a particular party whilst protest voters are those who are motivated by the aforementioned incentives behind casting a protest vote. The problem faced by protest voters in this setup can be divided into two components. Firstly, they need to understand the magnitude of the coordination problem that they face, that is, what the underlying distribution of preferences in the electorate is and, thereby, how many votes for the anti-mainstream are required to effect a successful protest. Secondly, they need to solve the identified problem by ensuring that, given the underlying core support for the anti-mainstream, the number of votes cast for the anti-mainstream falls within the critical region. Opinion polls may affect behaviour through both components. We therefore argue that there exist two channels through which the opinion poll may affect the decision to cast a protest vote, which we call the *information channel* and the *coordination channel*.

Motivated by theoretical predictions, we hypothesise that opinion polls increase the likelihood of a successful protest through both of these channels. Our experimental results provide support for these predictions. We also find evidence that polls increase the rate of protest voting and support for the idea that the polls enable subjects to update their beliefs about the true underlying distribution of preferences in the electorate in the right direction.

Our findings have interesting policy implications and may be relevant for those responsible for regulating the dissemination of information during election campaigns. For example, if those regulators take the view that voters should vote in line with their true underlying preferences, then they may wish to understand how the dissemination of information on voting intentions impacts the likelihood that individuals cast a protest vote. Some countries impose an embargo on the publication of pre-election polls for a certain amount of time preceding election day (Frankovic et al., 2018). This has been shown to increase the likelihood of coordination failures in terms of voters being less likely to cast an effective strategic vote and, thereby, being more likely to waste their vote (Lago et al., 2015). Our results suggest that such an embargo will also limit the extent of protest voting.

In the next section, we discuss the relevant literature and Section 3 presents our theoretical predictions. Section 4 describes the experimental design, we present our research hypotheses in Section 5, Section 6 presents our results and we conclude in Section 7.

# 2 State of the Art

There is an abundance of theoretical work aiming to explain voter behaviour that appears inconsistent with the assumption that voters cast their vote rationally to maximise their expected utility with respect to the outcome of the upcoming election. Riker & Ordeshook (1968) refine the expected utility approach and argue that voters obtain additional utility from phenomena such as affirming a partisan preference for a particular candidate, which are effectively independent of that candidate's prospects of winning the election. Similarly, Brennan & Hamlin (1998) set out a dichotomy between an instrumental account of voting, in which voters vote for the candidate that they perceive will leave them best off, and an expressive account, under which voters obtain utility from the kinds of phenomena proposed by Riker & Ordeshook (1968). Other papers highlight the importance of the role of information regarding candidate quality in the turnout decision (Feddersen & Pesendorfer, 1996; 1999). Such extensions have been used to explain historical puzzles in the literature, such as why voters turn out to vote at all when the expected probability of casting a pivotal vote is so small (Engelen, 2006; Blais, 2014).

Voters' party or candidate choice is often framed as being either *sincere*, in which a voter casts a ballot that directly reflects their true preferences, or *strategic* in which a voter deviates from voting for their true first preference. The latter may take the form, for example, of voting for a party with a more realistic chance of winning, in order to try to best influence the identity of the election winner (e.g. Fisher, 2004) or, it may reflect coordinating on a less-preferred party to avoid that a Condorcet loser wins the election (Tyszler & Schram, 2016).

Protest and strategic voting share some similarities; both involve casting an insincere vote, that is, a vote for a candidate inconsistent with a voter's first preferences (Franklin et al., 1994; Schimpf, 2021). However, the two behaviours are differentiated by the motives that underlie them; the strategic voter is motivated by influencing the election outcome directly, whereas the protest voter aims to send a targeted signal of dissatisfaction (Kselman & Niou, 2011). In addition, the anti-coordination features of protest voting also differentiate it from strategic voting – for example, voters with incentives to vote strategically to defeat a mutually disliked candidate gain by coordinating their votes behind a single challenger candidate, to avoid splitting the vote and allowing the disliked candidate to win (Tyszler & Schram, 2016; Myatt, 2017). This is arguably easier than coordinating on who amongst the like-minded voters should support the protest and who should not do so.

Protest voting is closely related to the literature on models that argue that aside from being used to effect the electoral outcome, voting can also have a signalling function. A voter's signal may be motivated by the intention of influencing the candidate's beliefs on the distribution of voter preferences, such that voters who do not vote for the winning candidate can still influence the winning candidate's policy offering (Razin, 2003). Other models focus on how intertemporal considerations affect voter behaviour; many voters focus not only on the outcome of the current election but are concerned with how votes in the current election can impact future elections. In these models, votes that perform a signalling function can be used as a means of influencing future policy platforms (Piketty, 2000; Castanheira, 2003). These models indicate that any model that considers only the direct benefits from the current election outcome may be ill-suited to understand the full range of voter behaviour.

One signal that voters may wish to send is that they are discontent with the policy offering of their most preferred party; this could be achieved by voting for a smaller party in the hope that the preferred party adopts some of the policy platform offered by that smaller party (Franklin et al., 1994). Such a vote can be understood within the aforementioned instrumental/expressive paradigm as an expressive tactical (or strategic) vote; voters obtain some expressive utility from the signal to the preferred party and are not motivated primarily by affecting which candidate wins the election. Alternatively, such a vote can be understood by any model that includes an intertemporal framework; the protest voter believes that a sufficiently large number of votes for the protest encourages the mainstream to change their policy (Myatt, 2017). The implementation of a revised policy, which is more closely aligned to the protest voter's preferences, will yield additional utility in future periods. As pointed out by Alvarez et al. (2018), this type of behaviour has been later defined as a tactical protest vote. In this paper, we tend towards the intertemporal interpretation of a protest vote. We capture this in a one-election model in which the additional utility provided by a successful protest is included in reduced form.

There exist a number of contributions that model tactical protest voting. Kang (2004) shows that protest voting occurs where voters have witnessed a decline in the quality of the major party that they would usually support, and a viable alternative is available. Kselman & Niou (2011) define the set of potential protest voters in a three-party election as those whose ideological positions lie between the leftmost and rightmost parties and show that the likelihood of protest voting decreases with the difference between the utility these voters derive from their first preference and the remaining two parties. Myatt (2017) provides a model that represents protest voting as a scenario in which anti-coordination is rewarded. A key insight in his study is that protest votes can be thought of as strategic substitutes; if a voter expects others to also engage in protest voting, that may reduce their incentives to do so because of the risk of the protest cause actually winning the election. There is therefore a non-monotonic relationship between the expected enthusiasm for a protest cause and the likelihood of casting a protest vote. This highlights a further distinction with more classical forms of strategic voting, under which such votes are strategic complements as an increased likelihood of the strategic cause winning the election drives the potential strategic voter to also support that cause. The Myatt model has been extended to incorporate interaction between this behaviour and financial markets (Wang & Zhou, 2023); in our paper we provide another extension, namely the effect of opinion polls on protest voting.

There is significant empirical evidence on the existence of the underlying logic of protest voting, that is, of voters switching to alternative options when the mainstream parties are considered to be of low quality (Schimpf, 2019), or where distrust or disillusionment in those parties is high (Bergh, 2004; Kang, 2004). Yet identifying a protest vote in actual election data can be a difficult task. Some papers (such as Bergh, 2004 and Schimpf, 2019) have sought to measure protest voting using survey approaches, but the capacity of this research to shed light on the strategic calculus undertaken by potential protest voters is limited. Empirical approaches may lack information on voters' true preferences and their expectations about the distribution of preferences within their electorate. Empirical studies based on observational field data also make it more difficult to isolate the impact of a specific institution like opinion polls on the likelihood of voters casting protest votes. To overcome some of these shortcomings, we employ an experimental approach to test how opinion polls impact the frequency of successful protest voting.

There has been much research on how opinion polls impact voter behaviour. Simon (1954) draws a distinction between bandwagon and underdog effects that characterise how voters respond to the information provided by an opinion poll. The bandwagon effect describes how voters may be more likely to vote for a candidate if the poll indicates that that candidate is likely to win the election; the underdog effect describes how voters may be more likely to vote for a candidate if the poll indicates that the candidate is unlikely to win the election. A large number of works have tested these effects and how they influence voter decisions on both turnout and candidate choice (e.g. Levine & Palfrey, 2007; Großer & Schram, 2010; Agranov et al., 2018; Boukouras et al., 2023). In a series of influential works in the field, Forsythe et al., 1993) show that polls play a significant role in facilitating coordination among voters with incentives to vote strategically. We contribute to this literature by studying the impact of polls in a setting where voters have a different incentive: that of casting a protest vote.

Opinion polls may reveal information on the distribution of preferences in the electorate and, where a multiplicity of equilibria exist, enable voters to coordinate their votes on a viable candidate to ensure that their vote is not wasted (Forsythe et al., 1993, 1996; Fey, 1997). Agranov et al. (2018) study the role of polls experimentally and use treatments that either perfectly reveal the distribution of voter types to subjects or enable subjects to participate in laboratory polls in which they declare their intentions, thereby studying both an environment where polls perfectly reveal information on the preference distribution and one under which subjects can respond strategically to polls, which may foster additional coordination (Andonie & Kuzmics, 2012).[6] In this paper, we also employ a laboratory study to research how the different functions provided by opinion polls allow for subjects to coordinate, but our focus is on their effect on protest voting.

Our experimental design used to model the scenario faced by protest voters bears similarities to a step-level public goods game with a binary contribution decision. Under these games, subjects are given an endowment and have the option to contribute that endowment to a public good. If a specific threshold is met, all subjects receive an additional payoff. These games have been tested in laboratory settings (for example in van de Kragt et al., 1983; Dawes et al., 1986; Offerman et al., 1996). We deviate from this classic setup by introducing a second threshold; relative to the optimal outcome (where the protest succeeds), subjects in our experiment risk losing money because either too few or too many people contribute to the public good (i.e., vote for the protest party). This second risk, that results in an electoral accident yields the lowest payoff to subjects in our experiment.

Overall, our paper is most closely related to Louis et al. (2022), which is the first experimental study that tests the logic of the type of protest voting modelled by Myatt (2017). Louis et al. (2022) focus on how protest voting is affected by the popularity of the protest and the protest's salience, that is, how much additional utility a successful protest brings. The project presented here applies insights from these various strands of the literature to focus instead on the specific role that pre-election opinion polls have on the likelihood of voters engaging in protest voting, and of those electoral protests being successful.

[6] Agranov et al. (2018) differs from our paper in that the authors are studying an environment with a more straightforward voting game where subjects are paid higher amounts where their preferred candidate wins and do not have protest-type preferences of the kind we study in this paper.

# 3 Theory

We consider an electorate consisting of $n$ voters (with $n > 1$ odd) that has to choose between a mainstream candidate $M$ and an anti-mainstream candidate $A$.[7] The winner of the election is determined by simple majority rule and abstention is not possible. There are three types of voters, which differ in the utility they obtain from different electoral outcomes. First, mainstream voters prefer candidate $M$ to win and the margin by which $M$ does so is irrelevant to them; they receive a (normalized) utility of 1 if $M$ wins and 0 otherwise. Second, for anti-mainstream voters it is exactly the other way around. They prefer $A$ to win and get a payoff 1 iff this happens. Finally, preferences of protest voter types not only depend on the winner of the election, but also on the candidates' vote shares. In particular, they prefer mainstream candidate $M$ to win, but at the same time candidate $A$ to get a sufficient number of votes to signal their discontent with the current policy stance of candidate $M$. We thus define their preferences in terms of two thresholds, $\underline{t}$ and $\bar{t}$, where $\underline{t} < \bar{t} = \frac{n-1}{2}$.[8] If the number of votes for $A$ (denoted by $t$) falls short of the lower threshold $\underline{t}$, the mainstream candidate $M$ wins and the protest fails; protest voters then get a normalized payoff of 1. If instead $t$ falls within the *critical range* $[\underline{t}, \bar{t}]$, $M$ still wins but the protest succeeds. Protest voters then receive a payoff equal to $1+s$; parameter $s > 0$ reflects the additional utility they obtain from the protest succeeding. Finally, in case the anti-mainstream party wins the election, i.e. if $t > \bar{t}$, protest voters obtain a payoff of zero. Protest voters thus have a joint goal that the electoral support for $A$ – from anti-mainstream voters and protest voters together – ends up in the critical range $[\underline{t}, \bar{t}]$.

Let $m$ denote the number of mainstream voters, $a$ the number of anti-mainstream voters, and $p$ the number of protest voters. Obviously, $m+a+p=n$. The type composition of the electorate is then reflected by vector $\mu = (m,a,p)$. In reality, voters do not know the exact type composition and polls may serve a useful purpose in providing more information about it. In line with this, also in our experiment the exact type distribution that prevails is unknown to the subjects. However, to build intuition, we will first consider the case in which $\mu$ is common knowledge.

Note that the decision problem for mainstream voters and anti-mainstream voters is trivial; they both have a weakly dominant strategy to vote for $M$ and $A$, respectively. Throughout we thus assume that they do so and the focus is on how protest voters vote. We let $\pi \in [0,1]$ denote the probability with which an individual protest voter casts a protest vote by voting for $A$. We restrict our attention to symmetric equilibria in weakly undominated strategies and denote these by $\pi^e$.

[7] In this section we provide an overview of our theoretical results. Please see Appendix A for all derivations.

[8] Note that under a simple majority rule, $A$ wins iff $t \geq \frac{n+1}{2}$ and thus iff $t > \frac{n+1}{2} - 1 = \frac{n-1}{2}$. This gives $\bar{t} = \frac{n-1}{2}$.

## 3.1 Full information benchmark

Suppose $\mu = (m, a, p)$ is common knowledge and consider the decision problem for a given protest voter indexed by $i$. Her vote matters only if it is pivotal in either achieving or abandoning the critical range $[\underline{t}, \bar{t}]$. Therefore, let $t_{-i}$ denote the number of votes cast for candidate $A$ among all $n-1$ other voters (irrespective of their type). The expected net benefit for $i$ of voting for $A$ rather than for $M$, i.e. of casting a *protest vote*, then equals:

$$NB_i\,(\text{protest vote}\,|\,\pi,\mu) = \Pr\left(t_{-i} = \underline{t} - 1\,|\,\pi,\mu\right) \cdot s - \Pr\left(t_{-i} = \bar{t}\,|\,\pi,\mu\right) \cdot (1+s) \qquad (1)$$

This expression can be easily understood. If $t_{-i} = \underline{t} - 1$, a protest vote from voter $i$ is necessarily needed to turn the protest into a success. In that case, $i$'s net benefit of voting for $A$ rather than for $M$ equals $1 + s - 1 = s$. $\Pr\left(t_{-i} = \underline{t} - 1\,|\,\pi,\mu\right)$ denotes the probability with which this eventuality happens, assuming all other protest voters cast a protest vote with probability $\pi$. In case $t_{-i} = \bar{t}$, a protest vote from $i$ will lead to $A$ winning the election, while a vote for $M$ would yield a mainstream victory together with the protest being successful. The net benefit of casting a protest vote is then $0 - (1+s) = -(1+s)$. Similar to before, $\Pr\left(t_{-i} = \bar{t}\,|\,\pi,\mu\right)$ reflects the probability of this eventuality occuring. For all other values of $t_{-i}$ the vote of voter $i$ is immaterial, as it neither affects the protest's success nor the electoral outcome.

Based on equation (1) four different scenarios can be distinguished, depending on whether or not each of the two pivotal probabilities can be positive for some value of $\pi \in [0,1]$. In the *indifferent scenario* both probabilities always equal zero, such that voter $i$ is never pivotal and therefore indifferent about whom to vote for. This necessarily happens when the electoral composition $\mu$ is such that the group of protest voters is too small to make a difference for both the protest's success and the electoral outcome. Because this scenario provides voter $i$ no guidance at all on how to vote, we can disregard those situations in the sequel. Three relevant scenarios thus remain.

In the *protest scenario* where only $\Pr\left(t_{-i} = \underline{t} - 1\,|\,\pi,\mu\right) > 0$ for some $\pi \in [0,1]$, the net benefit of casting a protest vote is (weakly) positive. Because the protest's success then possibly depends on $i$'s protest vote while the electoral outcome never does, voter $i$ has a pure protest motive and is best off casting a protest vote. This scenario applies when $a < \underline{t} \leq a + p \leq \bar{t}$. A unique Nash equilibrium in weakly undominated strategies then exists in which $\pi^e = 1$.[9] The opposite case arises when only $\Pr\left(t_{-i} = \bar{t}\,|\,\pi,\mu\right) > 0$ for some $\pi \in [0,1]$. It is then weakly dominant for $i$ to support the mainstream candidate, to avoid the risk of voting $M$ out of office by casting a protest vote. This *support scenario* allows a unique Nash equilibrium in weakly undominated strategies with $\pi^e = 0$. It occurs in electorates satisfying $\underline{t} \leq a \leq \bar{t} < a + p$.

[9]Note that a symmetric equilibrium in weakly dominated strategies with $\pi^e = 0$ exists side by side if $a < \underline{t} - 1$, because then more than one protest voter casting a protest vote is needed to make the protest a success. As noted earlier, we focus on equilibria in weakly undominated strategies.

Finally, in the *mixed scenario* a protest voter has mixed motives, because she can be pivotal in either way. This occurs when $a < \underline{t} < \bar{t} < a+p$. A mixed strategy equilibrium $\pi^e \in (0,1)$ then exists for which the net benefits of casting a protest vote as reflected in (1) equal zero. For the specific mixed scenario within our experimental parametrisation this reduces to:[10]

$$\pi^e_{mixed} = \frac{\sqrt{s}}{\sqrt{s}+\sqrt{1+s}} = \frac{1}{1+\sqrt{\frac{1}{s}+1}} \tag{2}$$

Intuitively, the higher the benefit $s$ from a successful protest, the more likely a protest voter is to cast a protest vote. Also note that $\pi^e_{mixed} < \frac{1}{2}$. This intuitively follows from the cost of 'overshooting' the critical range $(1+s)$ being larger for protest voters than the cost $s$ of 'undershooting' it (together with these two types of coordination failures being equally likely for $\pi = \frac{1}{2}$).

## 3.2 No information

Next suppose that the exact composition of the electorate is unknown to voters. A protest voter then has to form beliefs about the specific composition $\mu_j = \left(m_j, a_j, p_j\right)$ that applies (with $j \in \{1,...,J\}$ and $J$ the overall number of possible electoral compositions), knowing that she herself is a protest type. Let $\rho_j$ denote her prior belief that composition $\mu_j$ prevails and $\rho = (\rho_1,...,\rho_J)$ (with $0 \le \rho_j \le 1$ and $\sum_{j=1}^J \rho_j = 1$). Without further information, the cost-benefit analysis of casting a protest vote is then based on taking the $\rho_j$-weighted average of the expected net benefits of voting anti-mainstream over all the electoral compositions that may potentially apply:

$$NB_i\left(\text{protest vote}\,|\,\pi,\rho\right) = \sum_{j=1}^{J} \rho_j \cdot NB_i\left(\text{protest vote}\,|\,\pi,\mu_j\right) \tag{3}$$

In general many different electoral compositions may be possible, each one falling in one of the scenarios distinguished in the previous subsection. To simplify the exposition and in line with our experimental design, we focus on the simplest setting in which all three relevant scenarios are represented. That is, let $j \in \{1,2,3\}$, with composition $\mu_1$ corresponding to a protest motive scenario, $\mu_2$ to a mixed motive scenario, and $\mu_3$ to a support motive scenario. Moreover, these

[10] In Appendix A we provide a full characterisation of $\pi^e$ for the general case and show that the interior solution to $NB_i\left(\text{protest vote}\,|\,\pi,\mu\right) = 0$ is unique. Besides a mixed strategy equilibrium, also a pure strategy equilibrium $\pi^e = 0$ exists if $a < \underline{t} - 1$. Intuitively, then at least two protest votes from protest types are needed to make the protest a success, such that an individual protest voter cannot be pivotal if all other protest voters vote mainstream. This in turn makes voting mainstream a (weak) best response and thus $\pi^e = 0$ an equilibrium. Note that the latter equilibrium is not *responsive* (cf. Louis et al, 2022); protest voters employ the exact same strategy as mainstream voters. In the mixed scenario of our experimental design we have $a = \underline{t} - 1$ and hence no equilibrium with $\pi^e = 0$ exists. Similarly, when $m < \bar{t}$ in general also a non-responsive equilibrium with $\pi^e = 1$ exists side-by-side. As $m = \bar{t}$ in our experiment, also this equilibrium is excluded there.

specific compositions and their frequencies of occurrence are such that there is a unique mixed strategy equilibrium satisfying:

$$\pi_{NI}^{e}=\frac{\sqrt{\rho_2 s-\rho_3(1+s)}}{\sqrt{\rho_2 s-\rho_3(1+s)}+\sqrt{\rho_2(1+s)-\rho_1 s}}=\frac{1}{1+\sqrt{\frac{\rho_2(1+s)-\rho_1 s}{\rho_2 s-\rho_3(1+s)}}} \tag{4}$$

Note that $\pi_{NI}^{e}$ is increasing in $\rho_1$. Intuitively, if it becomes more likely that protest motive scenario $\mu_1$ applies, the equilibrium probability with which a protest voter casts a protest vote increases. Similarly, the higher the likelihood $\rho_3$ that support motive scenario $\mu_3$ applies, the lower $\pi_{NI}^{e}$ is. For $\rho_2=1$ (and thus $\rho_1=\rho_3=0$) equality (4) reduces to (2). More generally, if $\rho_1=\left[\frac{1+s}{s}\right]^2\rho_3$ – which holds in our experimental design – we have that $\pi_{NI}^{e}=\pi_{mixed}^{e}$. In that case the expected likelihood of casting a protest vote is necessarily higher under full information than under no information: $E\left[\pi_{Full}^{e}\right]=\rho_1\cdot 1+\rho_2\cdot\pi_{mixed}^{e}+\rho_3\cdot 0>\pi_{mixed}^{e}=\pi_{NI}^{e}$. The same holds for the expected likelihood that the protest is a success (and the expected utility a protest voter obtains in equilibrium).[11] This illustrates the general insight that more information increases the expected occurrence of protest voting, as well as the likelihood that the protest succeeds. It underlies the information channel effect of opinion polls.

## 3.3 Opinion polls: information channel

The outcome of opinion polls may provide voters with information about the composition of the electorate, allowing them to update their prior belief $\rho$. To capture this, suppose that, before the actual elections take place, some subset of voters is asked their (non-binding) voting intentions; the outcome of this poll becomes common knowledge. Let $\delta$ denote the fraction of declared intentions to vote for $A$ in the polls, corresponding to $d=\delta\cdot n$ intended $A$-votes in the overall electorate. A low $d$ indicates that the potential vote base for candidate $A$ is likely to be small, either because there are few anti-mainstream voters (small $a$), or there are few protest voters expected to cast a protest vote (captured by $p\cdot\pi$). How to interpret a particular polling outcome $d$ thus also depends on the likelihood $\sigma\in[0,1]$ with which individual protest voters declare an intention to vote for $A$ if they are polled, and the extent to which these poll declarations are representative of what they ultimately will do (i.e. how $\sigma$ compares to $\pi$).[12]

Based on the observed $d$ and the general tendency $\sigma$ of protest voters to poll in favour of $A$, voters can update their beliefs about the scenario that applies from $\rho_j$ to $\widehat{\rho}_j(d;\sigma)$ for all $j$. If polls are solely used in this way to guide protest voters' actual voting behaviour, the cost-

[11] This follows because in case mixed scenario $\mu_2$ applies, the likelihood of success is the same under full as under no information, given that $\pi_{NI}^{e}=\pi_{mixed}^{e}$. If either scenario $\mu_1$ or $\mu_3$ applies, the likelihood of a successful protest is strictly larger under full information. The expected utility of a protest voter equals: $1+\Pr(\text{success})\cdot s-\Pr(\text{overshoot})$.

[12] We assume that mainstream voters always declare an intention to vote for $M$ if they are polled and anti-mainstream voters declare an intention to vote for $A$.

benefit analysis equals the one under no information after replacing $\rho_j$ with $\widehat{\rho}_j(d;\sigma)$. The three electoral compositions used in our experiment are chosen such that a low $d$ provides conclusive evidence that protest scenario $\mu_1$ applies, whereas a high $d$ perfectly reveals support scenario $\mu_3$. For intermediate values of $d$ voters cannot draw such firm conclusions, but they can least exclude either the support scenario (when $d$ is low within the intermediate range) or the protest scenario ($d$ high within the intermediate range). Overall, it holds that the higher $d$, the less likely the protest scenario becomes and the more likely the support scenario: $\widehat{\rho}_1(d;\sigma)$ is decreasing in $d$ and $\widehat{\rho}_3(d;\sigma)$ increasing, with $\widehat{\rho}_2(d;\sigma)$ first increasing and then decreasing in $d$.

Drawing on the earlier general insights regarding the impact of more information, polls are expected to increase both the overall likelihood of protest voting and the protest becoming a success. Moreover, the rate of actual protest voting is expected to decrease with the support for the anti-mainstream party stated in the polls.

## 3.4 Opinion polls: coordination channel

Aside from providing information about the electoral composition, polls may also help protest voters in another useful way. To see this, note that even if protest voters were to know with certainty that mixed scenario $\mu_2$ applies, they would still face a difficult coordination problem. Instead of all of them individually voting according to mixed strategy $\pi^e_{mixed}$ in the actual elections, they would collectively be better off if they could coordinate their votes in some way such that the protest's success is secured. If the results from the polls indicate that the protest is very likely to succeed, i.e. if $d \in [\underline{t}, \bar{t}]$, protest voters have very little reason to deviate from their original intentions as reflected by the polls. If they indeed do not do so and the poll is representative, the actual election outcome will then very likely be in line with the poll. Effectively, the poll has given the protest voters an implicit way to coordinate within their own group.[13]

Theoretically we can capture this coordination channel in the following way. Instead of individual protest voters using a mixed strategy as in (4) based on updated beliefs $\widehat{\rho}_j(d;\sigma)$, in case $d \in [\underline{t}, \bar{t}]$ protest voters now as a group vote in such a way that the actual election outcome is fully in line with the poll outcome and the protest thus succeeds. Such implicit coordination makes it more likely that the protest succeeds, relative to the case where only the informational channel of polls is used (irrespective of how voters behave in the polls as captured by $\sigma$). This coordination channel is expected to be stronger the more voters can rely on the poll being representative of protest voters' final voting behaviour. In the experiment we capture this variation in the (recognised) representativeness of the polls by comparing the case of an exogenous poll, in which the protest voters under consideration did not take part, with the case of an endogenous poll where all protest voters under consideration do take part (and thus should realize the

[13]This 'copy-and-paste' strategy is successful if all voters know what they would declare in a poll if asked and remember this intention when the poll result is made known.

polls' perfect representativeness). Theoretically, endogenous polls are expected to increase the likelihood of protest voting – and especially of the protest becoming a success – even further than exogenous polls do.

A final theoretical consideration remains. Point predictions in the presence of polls clearly require the specification of $\sigma$. For our formal analysis of the endogenous polls case in Appendix A we assume that protest voters in the pre-election polling stage are fully rational and do an overall cost-benefit analysis of polling in favour of $A$ similar to (3), assuming that the polls' outcome subsequently affects actual voting behaviour along both the information and the coordination channels. For our specific parametrisation this leads to a unique equilibrium value $\sigma^e$ as part of an overall perfect Bayesian equilibrium, which is used to calculate the specific point predictions regarding voting behaviour and electoral outcomes in the presence of polls. It holds that $\sigma^e > \pi^e_{NI}$; protest voters are thus more likely to state an intention to cast a protest vote than they would actually do in the absence of additional information (i.e. when not learning the poll outcome). This follows from protest voters rationally anticipating the information and coordination benefits that polls bring when deciding on their poll declaration. Although this makes the formal analysis consistent in the sense that protest voters are fully rational in both the first polling stage and in the subsequent voting stage, this does not seem particularly realistic. Arguably more realistic assumptions might be either that $\sigma^e = \pi^e_{NI}$, i.e. voters are sincere and state their "true" voting intentions in the polls, or $\sigma^e < \pi^e_{NI}$ because protest voters might be reluctant to openly express support for an extreme party. Even though such different values of $\sigma$ would lead to different point predictions, they would leave our comparative statics predictions with respect to the incidence of protest voting and its electoral consequences unaffected.

# 4 Experimental Design

## 4.1 Main Experiment Design

The main sessions[14] of the experiment were conducted from July-October 2022 at CREED[15] at the University of Amsterdam with funding provided by the Amsterdam Center for Behavioral

[14] This study was preregistered at https://aspredicted.org/5ttd-t7gv.pdf. A pilot study for this study was conducted in Summer 2021 (Feltham et al., 2021). That study involved a design that focuses on the effects of providing opinion polls to voters for different informational environments, that is, how opinion polls increase the likelihood of the protest being successful in environments where voters know the true underlying distribution of preferences and those where they do not. Differences in outcomes between the two environments were intended to measure the information effect of opinion polls. Our results indicated that the opinion poll increased the likelihood of a successful protest in the No information environment but not in the Full information environment (where evidence was mixed), although this treatment may have been subject to ceiling effects. Our previous pilot study did not test our theoretical predictions for the effect of the opinion poll. For this paper, we have revised the design to correspond closely to the theory. This allows us to more cleanly differentiate between the information and coordination channels.

[15] Center for Research in Experimental Economics and political Decision making.

Change. Subjects were recruited from the student population at the University of Amsterdam and paid approximately €20 on average for sessions lasting around 75 minutes. We collect data from 360 subjects, equally divided across three treatments.

Subjects are randomly sorted into one of the three treatments: *Control*, *Exogenous* Polls and *Endogenous* Polls in a between-subject design. All subjects are assigned the role of protest voters and take part in a series of mock elections. There are two candidates in the election, M and A, representing the mainstream and anti-mainstream candidates.[16] There is no option to abstain so subjects are required to cast a vote for one of the two candidates in each round. The electorate for each election consists of nine voters in total: three protest voters and six computerised core supporters who are programmed to play the weakly dominant strategies of a- and m-types, that is, to vote for the A and M candidates, respectively. Subjects are made aware of the total size of the electorate, the number of programmed voters, and the strategies of the programmed voters. We refer to the programmed voters as A- and M-supporters. Subjects are provided with a set of instructions (these can be found in Appendix B) and have to pass a series of attention checks to proceed with the experiment.

We operate matching groups of twelve subjects, allowing for four electorates of nine voters each (six of which are automated) per matching group in each round of the experiment. Between rounds, (protest) voters are re-matched within matching groups. After five practice rounds, subjects play 20 main rounds of the experiment. We collect data from ten matching groups per treatment, yielding ten statistically independent observations per treatment. This design enables us to collect $12 \times 10 \times 20 = 2,400$ individual-level observations per treatment and $4 \times 10 \times 20 = 800$ election-level observations per treatment.

Table 1: Different outcomes of election for parameters used in experiment

| **Votes cast for A-candidate** | **Winner of election** | **Successful protest?** | **Experimental points earned** |
|---|---|---|---|
| {0, 1, 2} | M | No | 10 |
| {3, 4} | M | Yes | 20 |
| {5, 6, 7, 8, 9} | A | No | 0 |

Notes: Depending on the number of votes cast for the anti-mainstream party (A), given in the first column, the remaining columns depict the winning party, whether or not the protest was successful and the earnings for a protest voter that result from these.

In the experiment, the critical threshold for a successful protest is defined as $\underline{t} = 3$ and the number of votes required for a candidate to win is 5, that is, $\bar{t} = 4$. The additional utility for a successful protest is set at $s = 1$ and voter payoffs in the experiment are multiplied by ten. The resulting payoff structure and outcome per round are summarised in Table 1. Recall that the votes cast for the *A*-candidate are summed across automated votes by core *A* supporters and protest votes cast by experimental subjects.

[16] These are labelled as A and B for the subjects.

Table 2: Experimental scenarios

| Scenario | (a, m, p) | Equilibrium | Motive | Frequency |
|---|---|---|---|---|
| 1 | (0, 6, 3) | Vote A | *Protest* | 0.4 |
| 2 | (2, 4, 3) | Mixed | *Mixed* | 0.5 |
| 3 | (4, 2, 3) | Vote M | *Support* | 0.1 |

Notes: The scenarios describe distinct electorate compositions. In each scenario there are three protest voters. The column 'Equilibrium' describes a protest voter's equilibrium vote, derived using the theory in Section 3.

To create uncertainty with respect to the composition of the electorate, we vary this randomly across rounds. In particular, for each election, the number of M- and A-supporters is determined as a random draw from one of three scenarios set out in Table 2. The probability of each scenario is given in the final column. Subjects are told the probability with which each of the three scenarios will occur at the start of the experiment, but not which scenario actually applies in any given round. The scenarios are designed to test a range of coordination problems underlying achieving a successful protest. The scenarios correspond to those distinguished between in Section 3. Scenario 1 is the *Protest scenario*. There are no anti-mainstream supporters. As a consequence, all three protest voters must vote for $A$ to realise the critical range. Voting for $A$ is then a weakly dominant strategy. At the other end, scenario 3 is the *Support scenario*; it involves four core supporters voting for $A$, so that any additional vote would cause an electoral accident. It is then a weakly dominant strategy for protest voters to vote for the mainstream, $M$. Finally, in the *Mixed scenario* 2, there are two core supporters in favour of $A$, so that either one or two out of the three protest voters must vote for $A$ to realise the critical range. The symmetric equilibrium then involves a mixed strategy.

We now turn to our poll treatments. In the *Endogenous* treatment, each round has two stages. In the first stage, subjects are asked which candidate that they intend to vote for. At the start of the experiment, they are informed that there is no requirement to vote for the same candidate in the election. Subjects are also told that programmed supporters automatically declare an intention for the candidate they support. The vote intentions of programmed voters and experimental subjects as stated in the poll are aggregated at the electorate-level and made available to all subjects in this treatment before they cast a vote in the election stage.

In the *Exogenous* treatment, each round has an additional stage in which the poll result taken from a corresponding *Endogenous* session with the same realised scenario is shown to subjects. After observing the poll result, subjects then cast their vote in that round.

In our theoretical analysis, a voter's belief about the realised distribution of types (i.e., the scenario) plays a key role. For this reason, we are interested in how subjects update their beliefs about the realised scenario in response to the information provided by the opinion poll. For this purpose, we elicit subjects' beliefs about which scenario applies using a binarised scoring rule

(Hossain & Okui, 2013).[17] Note that any given poll result rules out one possible scenario. In particular, if the polls indicate fewer than four votes for $A$, scenario 3 is ruled out because in that scenario all four core $A$-supporters would poll for $A$. Similarly, a poll indicating four or more votes for $A$ rules out scenario 1, because there must be at least one core $A$-supporter. To simplify the task for subjects, we eliminate the scenario that cannot theoretically occur on the basis of the observed poll result, and ask subjects only to provide their subjective belief (as a probability) that scenario 2 is the true state of the world.[18] In this way, subjects only have to provide one probability, rather than a distribution across the three scenarios. We elicit subjects' beliefs before they cast their vote in each round and, for the *Exogenous* and *Endogenous* treatments, after they observe the poll outcome. The beliefs task is run in every other round of the experiment.[19]

After the 20 rounds have been completed we use the Gneezy-Potters method to measure risk attitudes (Gneezy & Potters, 1997 and Charness & Gneezy, 2010 in Charness et al., 2020). Subjects are asked to invest any amount $x$ out of an endowment of \$8 (to the nearest euro cent), in a lottery that pays out $2.5x$ or 0 with equal probability. Note that risk aversion is decreasing in $x$.

After completion of the risk attitude elicitation, subjects are asked to fill out a short survey including questions about age, gender identity, educational background and how they made their decisions during the experiment. After this has been completed, participants are paid and dismissed. Aside from an €8 participation fee and the amount earned in the risk-elicitation task, earnings are determined in experimental points. We apply an exchange rate of €1 for each 10 experimental points earned across six randomly selected decision rounds and three randomly selected belief elicitations.

# 5 Hypotheses

## 5.1 The incidence of protest voting

As set out in more detail in Appendix A, our theory enables us to predict the frequency with which subjects cast a protest vote, that is, vote for the anti-mainstream candidate. We label this probability $\pi$. Our first set of hypotheses relates to comparative statics of the frequency of

[17] We apply the method as described by Wilson & Vespa (2018).

[18] The complement of the probability provided by subjects defines their belief that the other possible scenario was realised.

[19] Our application of the binarised scoring rule involves drawing two random numbers between 0 and 100. If scenario 2 applies and the stated probability is larger than either of these numbers, the subject earns ten experimental points. If scenario 2 does not apply and the stated probability is smaller than either of these numbers, the subject earns ten experimental points. In other cases, the subject earns nothing. It is explained to the subjects that their expected earnings are highest if they report their true beliefs. See the instructions in Appendix B for more details.

protest votes across treatments. For the studied case where the scenario is unknown to voters we calculate for each treatment the equilibrium $\pi^e$ and the corresponding (distribution of) equilibrium outcomes and take the weighted average over the three scenarios (using the frequency distribution in Table 2). Our theoretical predictions broken down by the value of $d$ can be seen in Table D1 in Appendix D, whilst the key aggregate predictions that motivate the hypotheses are presented in Table 3 below.

Table 3: Main theoretical predictions

| Treatment | Pr(protest vote) | Pr(undershoot) | Pr(success) | Pr(overshoot) |
|---|---|---|---|---|
| Control | 0.414 | 0.472 | 0.412 | 0.115 |
| Exogenous | 0.600 | 0.178 | 0.719 | 0.104 |
| Endogenous | 0.631 | 0.078 | 0.893 | 0.029 |

Notes: Cells denote the equilibrium probabilities of the events depicted in the column heads. These are based on the theory described in Section 3 and Appendix A.

Based on these predictions, we present our hypotheses relating to the rate of protest voting:

***Hypothesis 1a:*** *Protest voting is more likely under the Exogenous treatment than under the Control treatment.*
***Hypothesis 1b:*** *Protest voting is more likely under the Endogenous treatment than under the Exogenous treatment.*

The intuition behind these hypotheses follows from the theoretical discussion in Section 3. As explained there, more information in general yields increased protest voting. Moreover, besides additional information, only endogenous polls also provide opportunities for implicit coordination. Hypotheses 1a and 1b follow directly from these general insights.

## 5.2 Electorate-level outcomes

Table 3 also shows our theoretical predictions for the frequency of outcomes at the electorate level. The aggregate number of protest votes may: run short of the critical range, which we call '*undershooting*'; fall within this range, labelled '*success*', or may cause an electoral accident, labelled '*overshooting*'. The probabilities of these outcomes in equilibrium are shown in the final three columns of Table 3. This yields our next set of hypotheses:

***Hypothesis 2a:*** *The protest is more likely to succeed under the Exogenous treatment than under the Control treatment.*
***Hypothesis 2b:*** *The protest is more likely to succeed under the Endogenous treatment than under the Exogenous treatment.*

***Hypothesis 2c:*** *The difference between the occurrence of undershooting and the occurrence of overshooting is smaller under the Exogenous treatment than under the Control treatment.*
***Hypothesis 2d:*** *The difference between the occurrence of undershooting and the occurrence of overshooting is smaller under the Endogenous treatment than under the Exogenous treatment.*

Hypotheses 2a and 2b directly relate to the two channels we are interested in. Hypothesis 2a relates to the *information channel* – if we confirm this hypothesis then that represents evidence that the poll increases the likelihood of the protest succeeding through the information inherent in the value of $d$ that enables subjects to draw some inferences about the realised distribution of types in the electorate. If we confirm 2b, this is evidence in favour of the *coordination channel* – the poll also increases the likelihood of the protest succeeding by giving subjects an opportunity to condition their voting behaviour directly on the poll result by providing an additional shot at solving the coordination problem. Finally, the intuition underlying 2c and 2d is that the larger costs of overshooting compared to undershooting make protest voters reluctant to cast a protest vote and cause undershooting to occur more often. When information increases, protest voters become less reluctant as they can cast their protest vote more effectively. This especially mitigates undershooting while avoiding more overshooting at the same time. In fact, also overshooting is reduced, albeit to a smaller extent. The improvement in success probability that the information obtained from the polls brings is therefore to a larger extent due to less undershooting than to reduced overshooting.[20]

## 5.3 Mechanisms

We have argued that opinion polls can affect the decision to cast a protest vote by allowing voters to update their beliefs about the realised distribution of voter types (the information channel) and by providing a tool for direct coordination. Here, we provide hypotheses that allow us to directly test these two mechanisms.

### 5.3.1 The information channel: beliefs

We consider how subjects update their beliefs about the true state of the world in response to the information provided by the poll. Our theoretical predictions correspond to a posterior belief about the likelihood that scenario 2 is the true state of the world. As before, we denote this updated belief as $\widehat{\rho}_2$.

A first thing to note is that there are poll outcomes that perfectly reveal that $\widehat{\rho}_2 = 0$. For

[20]Note that $u_{Exo} - o_{Exo} < u_{Control} - o_{Control} \iff o_{Control} - o_{Exo} < u_{Control} - u_{Exo}$, where $u$ and $o$ denote the probabilities of undershooting and overshooting, respectively (and the subscripts refer to the treatment). Hypothesis 2c (and similarly 2d) is thus equivalent to this alternative formulation.

example, if $d = 0$ or $d = 1$, then a perfectly rational voter knows with certainty that scenario 1 applies ($a = 0$), so that $\widehat{\rho}_2 = 0$. Similarly, if $d = 6$ or $d = 7$, then this voter knows with certainty that scenario 3 applies ($a = 5$) because there are only three protest voters so that $a \geq 3$ necessarily. Again, $\widehat{\rho}_2 = 0$.[21] Our next hypothesis allows for not all voters being perfectly rational and predicts the weaker comparative static that protest voters' beliefs respond to poll results that imply $\widehat{\rho}_2 = 0$.

***Hypothesis 3:*** *For the Exogenous and Endogenous treatments, subjects are less likely to believe that the mixed scenario applies when the poll result perfectly reveals that* $\widehat{\rho}_2 = 0$ *(i.e. when* $d \in \{0, 1, 6, 7\}$*) than when it does not.*

Our next hypothesis follows directly from the discussion in Subsection 3.3:

***Hypothesis 4:*** *In the Exogenous and Endogenous treatments, the rate of protest voting weakly decreases in the value of d (in Endogenous conditional on the own polling statement,* $d_i$*).*

This hypothesis is motivated by the notion that, for low levels of $d$ (i.e. $d \leq 1$), subjects may realise that the protest scenario applies with certainty so that their weakly dominant strategy is to so cast a protest vote with probability 1. For $d \geq 2$, higher values of $d$ are associated with a higher risk of overshooting, which thereby decreases the equilibrium probability of casting a protest vote.

#### 5.3.2 Coordination Channel: Copy-and-Paste

The coordination channel allows protest voters to use the poll outcome as an indicator of who should cast a protest vote and who should not when there is a positive probability of the mixed scenario having been realised ($\widehat{\rho}_2 > 0$, i.e. when $d \in \{2, 3, 4, 5\}$). If this is the case and the poll result falls within the critical range ($d \in \{3, 4\}$), then a straightforward way to ensure a successful protest would be for all protest voters to vote for the party that they declared in the poll. To allow for noise, our hypothesis predicts a weaker comparative static than this perfect copy-and-paste.[22] In particular, we hypothesise that a copy-and-paste strategy is most often used when it makes most sense to do so.

***Hypothesis 5:*** *In the Endogenous treatment, protest voters are more likely to use a copy-and-paste strategy (i.e. vote in line with their declared intentions in the poll) when the poll outcome*

[21] In addition, if a rational protest voter declared in the poll that she would vote for $M$, then she knows for $d = 5$ that (because there are only two other protest voters) $a \geq 3$, and therefore $\widehat{\rho}_2 = 0$. Similarly, a protest voter that polled for A and observes $d = 2$ as aggregate polls outcome can conclude that $\widehat{\rho}_1 = 1$ and thus $\widehat{\rho}_2 = 0$.

[22] Within our experimental setting, such noise can occur because not all participants are perfectly rational. Outside the context of our experiment, noise can occur, for example, because not all voters are polled.

*falls within the critical range {3, 4} than when it does not.*

# 6 Results

To test our hypotheses, we conduct non-parametric statistical tests. In particular, we run Fisher-Pitman permutation tests with 10,000 permutations, taking the matching group as the unit of observation. Within each matching group and depending on the outcome of interest, we average across individuals or electorates such that each permutation test involves comparing two (or more) subsamples of ten statistically independent observations. Our key results, which are referred to and presented in this section, can be found in full detail in Table D2 in Appendix D.

## 6.1 The Incidence of Protest Voting

We start with considering the frequency of protest voting. Our focus here is on the protest votes by our subjects as a fraction of their votes cast, aggregated across rounds (and excluding the practice rounds).[23] Figure 1 shows this fraction across treatments. Casual inspection shows that protest voting increases as we move from Control to Exogenous and from Exogenous to Endogenous. Moreover, in the treatments with polls, there appears to be less protest voting than theoretically predicted.

The latter conclusion is easily confirmed in Figure 1; contrary to the Control treatment, the theoretical predictions do not fall within the 95% confidence interval of the observed votes. Our Fisher-Pitman tests confirm that the observed treatment differences are highly significant. This provides direct evidence in favour of Hypotheses 1a and 1b, giving:

***Result 1a:*** *there is more protest voting when there are exogenous polls than without polls.*
***Result 1b:*** *there is more protest voting when there are endogenous polls than when there are exogenous polls.*

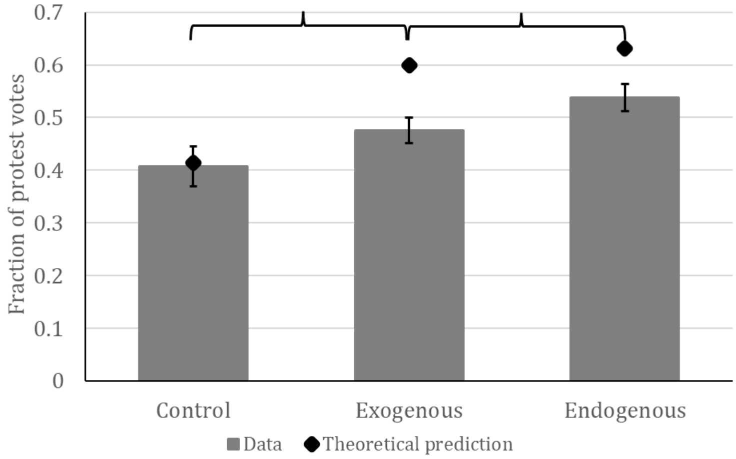
These results point to a role for both the information and coordination effects. We return to this issue in the next two subsections.

## 6.2 Electorate-level Outcomes

We now turn to the results at the electorate level. Once again, our focus here is on the outcome aggregated across rounds. Figure 2 shows the distribution of electoral outcomes across un-

[23] In appendix C we show this fraction across rounds (cf. Figure C1). This shows that (1) behaviour stabilises after the practice rounds; (2) the theory is relatively effective in predicting behaviour, especially for the predictions that do not lie at the extremes of the $[0,1]$ interval. For these predictions (i.e. those that relate to the $a = 0$ and $a = 4$ scenarios), subjects do not follow the predicted weakly dominant strategy as frequently as theorised.

Figure 1: Rate of protest voting




Notes: Bars show the fraction of protest votes per treatment. Because our unit of observation is the matching group, we first take the average per matching group and then across matching groups. Error bars show the 95% confidence intervals derived in this way. Significance levels are based on Fisher-Pitman permutation tests for this unit of observation.

dershooting, successful protest, and overshooting. Eyeballing the figure shows that successful protesting increases as we move from Control to Exogenous and from the latter to Endogenous. For successful protesting, the data for the treatments for polls once again fall short of the theoretical predictions. The figure also shows that undershooting is more prevalent than overshooting.[24]

The results of the Fisher-Pitman tests depicted in Figure 2 provide direct tests of our Hypotheses 2a-2d. This gives:

***Result 2a:*** *Protest voting is more often successful when there are exogenous polls than without polls.*

***Result 2b:*** *Protest voting is more often successful when there are endogenous polls than with exogenous polls.*

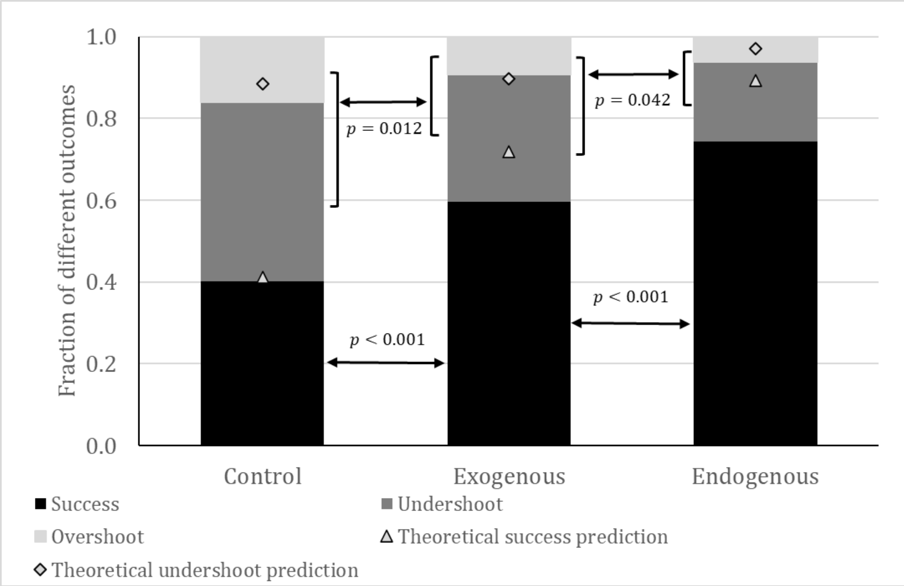
***Result 2c:*** *The difference between the occurrence of undershooting and the occurrence of overshooting is smaller when there are exogenous polls than without polls.*

***Result 2d:*** *The difference between the occurrence of undershooting and the occurrence of overshooting is smaller when there are endogenous polls than exogenous polls.*

These results confirm *Hypotheses 2*. The differences between the means concerned are all significant at at least the 5% level. This indicates that the poll plays both an informational role

[24]Once again, we refer to Appendix C for the results per round. Figure C2 shows the frequency with which the different outcomes occur across rounds. The main takeaways are (1) that the theory performs relatively well in predicting successful protest across rounds in the *Control* treatment. (2) For the other two treatments, successful protest tends to fall short of the theoretical predictions.

Figure 2: Vote outcomes




Notes: Bars show the distribution of electoral outcomes per treatment. Because our unit of observation is the matching group, we first take the average per matching group and then across matching groups. The lower p-values refer to the difference in success rates across treatments and are based on Fisher-Pitman permutation tests for this unit of observation. The upper p-values are based on the same test and compare the (difference in the) difference between undershooting and overshooting rates across treatments

(measured by the difference between the *Exogenous* and *Control* treatments) and a coordinating role (the difference between the *Endogenous* and *Exogenous* treatments). Though we reached the same conclusion when considering the incidence of protest voting in Section 6.1, it is important to note that an increase in protest votes does not necessarily imply increased success (it may also lead to more overshooting). Finally, note that our results also show that, as the rate of successful protesting increases, the rates of undershooting and overshooting both decrease, with undershooting falling to a larger extent (as predicted theoretically, cf. Table 3).

To shed more light on how the *Endogenous* treatment provides subjects with an opportunity to coordinate and thereby increase the likelihood that the protest succeeds, we present empirical transition matrices for the *Exogenous* and *Endogenous* treatments in Table 4. This table shows, for each poll outcome, the distribution of observations across the three possible electoral results. The fractions in each row therefore sum up to one. For example, a matrix with ones on the diagonal and zeros elsewhere would indicate that the polls perfectly predict election outcomes.

A notable difference between these two treatments is that the rate of conversion from a successful poll result to a success in the vote is much lower in the *Exogenous* treatment than in the *Endogenous* treatment. When testing for statistical differences in this rate between treatments (i.e. comparing the central cell of the *Exogenous* panel with the central cell of the *Endogenous* panel), we find that the Fisher-Pitman p-value is $< 0.001$, indicating that subjects in the *Endogenous* treatment convert a successful poll result to a successful protest in the election

Table 4: Transition matrix for *Exogenous* and *Endogenous* treatments.

| | | | | |
|---|---|---|---|---|
| Exogenous | | | | |
| | | **Actual outcome** | | |
| | | Undershoot | Succeed | Overshoot |
| **Outcome of poll (Hypothetical outcome)** | Undershoot | 0.46 | 0.49 | 0.06 |
| | Succeed | 0.23 | $0.68^{***}$ | 0.09 |
| | Overshoot | 0.09 | 0.69 | 0.22 |
| Endogenous | | | | |
| | | **Actual outcome** | | |
| | | Undershoot | Succeed | Overshoot |
| **Outcome of poll (Hypothetical outcome)** | Undershoot | 0.40 | 0.58 | 0.03 |
| | Succeed | 0.04 | $0.92^{***}$ | 0.04 |
| | Overshoot | 0.06 | 0.67 | 0.27 |

Notes: Each element of the table tells us the proportion of observations for which the combination of each poll outcome and actual outcome occurred. Data on the diagonal shows the frequency of entries for which the poll result coincided with the vote outcome. Stars represent the significance of p-values of Fisher-Pitman tests for tests of statistical difference between *Exogenous* and *Endogenous* treatments for conversion rate from successful protest in the poll result to successful protest in the election stage.

stage more effectively than subjects in the *Exogenous* treatment. We take this as further evidence of the coordination effect of polls.

## 6.3 Mechanisms

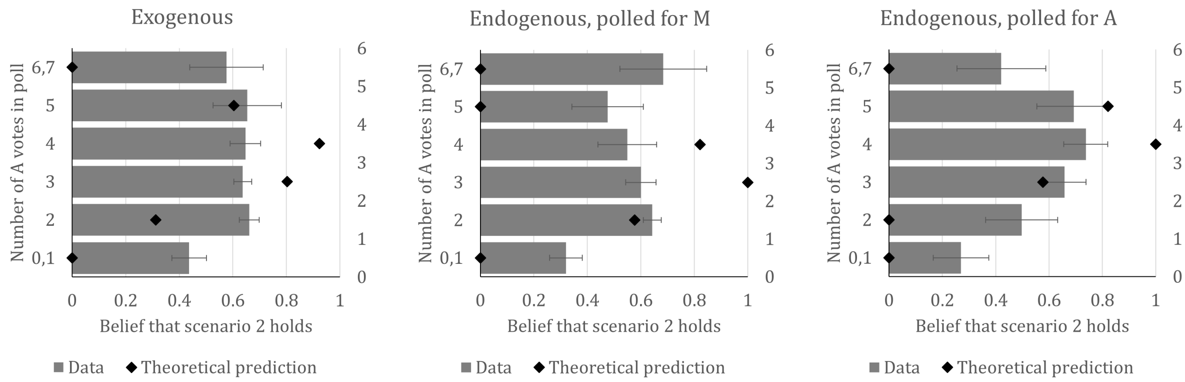
### 6.3.1 The information channel: beliefs

To start, Figure 3 plots the average beliefs for each poll outcome (as measured by $d$). We plot both our theoretical predictions and the data. For the *Endogenous* treatment, we plot separate series for subjects who themselves cast a poll for the mainstream candidate (M) and those who cast a poll for the anti-mainstream candidate (A). This is because, in this treatment and for the low number of voters in our experiment, the subject's own polling decision enables them to perfectly determine the true state of the world also for some intermediate poll outcomes; see footnote 21.

Recall that Hypothesis 3 predicts that protest voters respond to the information in the polls and are less likely to believe scenario 2 applies when the poll result perfectly reveals that it does not. To test this, we first pool the cases where the poll perfectly reveals $\rho_2 = 0$ and those that do not.[25] Figure 4 shows the average beliefs for the data pooled in this way. Though we observe a high predicted probability of scenario 2 having been realised even if the poll results indicate that this is not possible, the comparative statics are as predicted. The Fisher-Pitman tests we apply show that this effect is highly significant for both exogenous and endogenous polls.

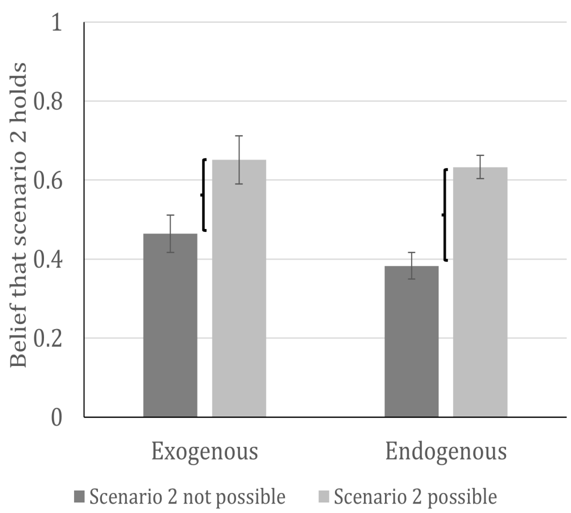
[25] Note from Figure 3 that with one exception, when the poll does not reveal $\rho_2 = 0$, the correctly updated belief is that $\rho_2 > 0.5$.

Figure 3: Beliefs after Polls




Notes: Bars show the reported beliefs that scenario 2 applies ($\rho_2$) for the poll result depicted on the vertical axis. Because our unit of observation is the matching group, we first take the average belief per matching group and then across matching groups. Error bars give 95% confidence intervals. Diamonds show the beliefs that follow from perfect Bayesian updating, given the equilibrium probability of polling for $A$ ($\sigma = 0.489$).

Figure 4: Beliefs for pooled data



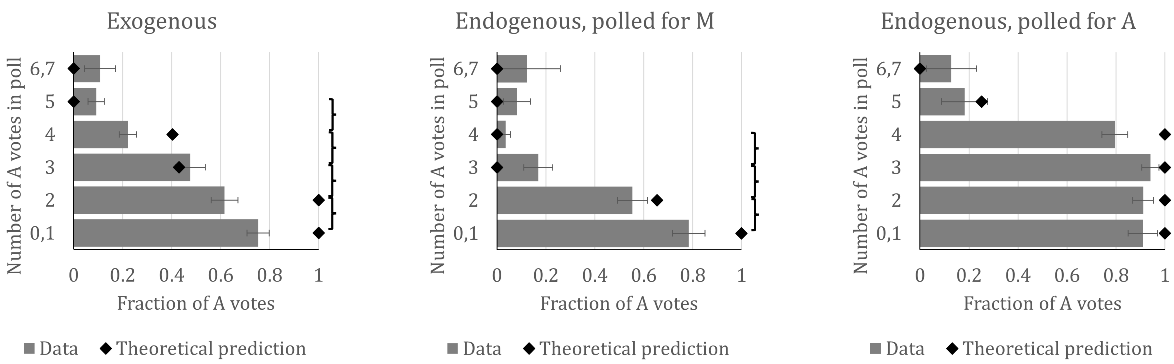

Notes: Bars show the reported beliefs that scenario 2 applies ($\rho_2$), pooled across cases where the poll result implies that $\rho_2 = 0$ and those that imply $\rho_2 > 0$. For the latter case, we pool the results where the voter polled for $M$ and $A$. We first take the average belief per matching group and then across matching groups. Error bars give 95% confidence intervals. The p-values refer to Fisher-Pitman tests for differences between the two scenarios.

This yields our next result:

***Result 3:*** *Protest voters' beliefs that scenario 2 has been realised is higher after polls reveal that this scenario is possible than when they reveal that it cannot have occurred.*

Moving from beliefs to choices, we argue in *Hypothesis 4* that the information provided by opinion polls will affect individual voting behaviour; more votes for *A* in the polls will lead to less protest voting. Figure 5 plots the average fraction of protest votes per matching group against the poll result.

Figure 5: Protest voting after Polls




Notes: Bars show the fraction of protest voters who voted for *A*, after the poll result depicted on the vertical axis. Because our unit of observation is the matching group, we first take the average protest votes per matching group and then across matching groups. Error bars give 95% confidence intervals. Brackets to the right of a graph indicate that a Fisher-Pitman test for the pairwise-difference between the poll outcomes concerned shows statistical significance at the 5% level or better. Details are available upon request.

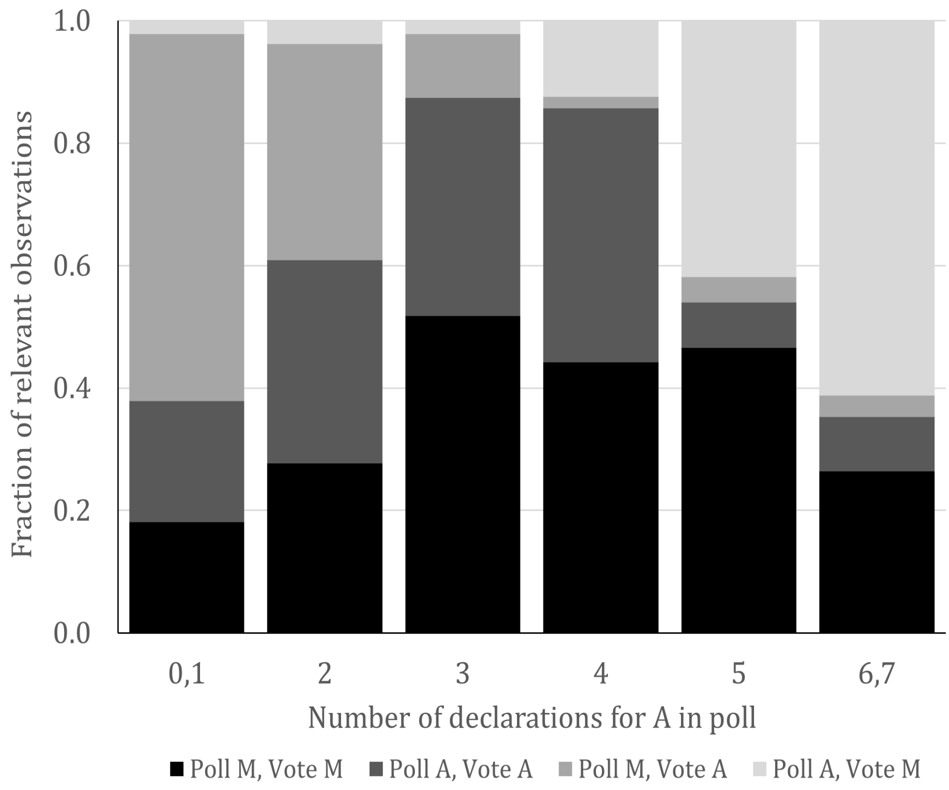
The figure shows that protest voting tends to decline with the poll result in all treatments. In the treatment with endogenous polls, however, it does appear to matter whether the voter opted for *M* or *A* in the poll. In the latter case, she is much more likely to vote for *A* in the election. This suggests a tendency towards consistency in the two decisions (further explored in the next subsection). Note that all significant differences between two adjacent poll results reflect a decline in protest voting for the higher poll result. Moreover, each graph shows at least two such significant decreases. Together, this provides support for Hypothesis 4:

***Result 4:*** *Protest voting is weakly decreasing in the number of votes for A in the polls (in Endogenous conditional on the own polling statement, $d_i$).*

### 6.3.2 The coordination channel: copy-and-paste

It is interesting to compare the two panels in Figure 5 for the *Endogenous* treatment. When the poll indicates $d \in \{3,4\}$ and thus predicts a successful protest, protest voters to a very large extent vote according to their declaration in the poll. That is, they consistently use a 'copy-and-paste' strategy. Outside of the critical range this is (much) less the case. In particular, when the poll indicates $d \in \{0,1\}$ and thus undershooting to be the likely outcome, protest voters predominantly vote for A in the actual elections, also those who initially polled for M. The latter protest voters (which in this case form a majority among all protest voters) thus use an 'inconsistent' voter strategy, where the polling choice differs from the ultimate voting decision. Similarly, in case $d \in \{6,7\}$, such that the poll clearly suggests an overshooting outcome, most protest voters vote for M in the elections. Those protest voters who initially polled for A (again the majority) then use an inconsistent voter strategy. For polling outcomes just outside the critical range (i.e. $d \in \{2,5\}$), voting behaviour is less pronounced. But also then the protest voters whose poll declaration contributed to being on the 'wrong side' of the relevant threshold, are quite likely to deviate from their initial voting intentions.

Figure 6: Voter strategies




Notes: Bars show the fraction of protest voters who employed different (Poll, Vote) strategies with inconsistent strategies stacked on top of consistent strategies.

To formally evaluate our Hypothesis 5, we plot in Figure 6 the frequency with which the four available voter strategies are used depending on the poll outcome. The two darker shaded areas at the lower end of the bars reflect the two consistent strategies *(Poll M, Vote M)* and *(Poll A, Vote A)*. The lighter shaded areas at the upper end of the bars reflect the inconsistent strategies *(Poll M, Vote A)* and *(Poll A, Vote M)*. The figure clearly shows that in the far majority of cases, viz. 86.7% of the time, subjects play a consistent strategy when $d \in \{3,4\}$.[26] This suggests that subjects understand that playing a copy-and-paste strategy can enable them to ensure a successful protest. Nevertheless, there is a statistically significant fraction of instances in which protest voters do not play a consistent strategy after a poll result in the critical range; a Fisher-Pitman test of the hypothesis that the fraction of consistent choices is equal to 1 is rejected ($p = 0.002$).

For poll results outside of the critical range (i.e. $d \notin \{3,4\}$), far fewer consistent strategies are used. Overall subjects then use a consistent strategy 57.7% of the time (with the highest fraction being 60.9% when $d = 2$). The Fisher-Pitman test confirms that this percentage is significantly different from the 86.7% consistency rate observed for poll results within the critical range ($p = 0.002$). This provides clear evidence that the copy-and-paste strategy that underlies the coordination effect of opinion polls is recognised and effectively applied by our participants, in line with our Hypothesis 5:

***Result 5:*** *Protest voters in the Endogenous treatment are more likely to use a copy-and-paste strategy when the poll outcome falls within the critical range* $\{3,4\}$ *than when it does not.*

# 7 Conclusion

In this paper, we study how opinion polls impact on the choice to cast a protest vote and on the frequency with which protest votes are successful. We argue that the opinion poll serves two functions: to provide information to voters with information on the distribution of preferences in the electorate and to act as a device that can help to solve the anti-coordination problem that arises amongst protest voters. Our experimental design allows us to determine the impact of both of these effects on our main outcomes of interest: the decision to cast a protest vote and whether or not the protest succeeds.

Our analysis is founded on a theoretical model that predicts two effects of opinion polls on protest voting. The *information effect* reflects the information that polls carry about the distribution of preferences in the electorate. The *coordination effect* highlights how polls help protest voters to overcome the anti-coordination problem that they face. Our experimental design allows us to first isolate the information effect by using polls (in the exogenous treatment) that provide information only about the underlying preference distribution and subsequently adding the coordination possibility in our endogenous treatment. The experimental results provide strong evidence of both effects. We find that both the rate of individual protest voting and the success of protest voting are higher in the exogenous polls treatment than in the control without polls. This suggests that polls provide subjects with information about how to more

[26] Once again, to obtain this percentage, we first determined the means within matching groups and then across matching groups. This is because we use the matching group as the unit of observation for our statistical tests.

effectively cast a protest vote. We also observe more protest voting and higher success in the endogenous poll treatment than in the exogenous case. We attribute the latter differences to the coordination effect. We observe that both the extent and the success of protest voting is higher with endogenous polls.

In order to better understand the mechanisms underlying the two effects, we first analyse subject's beliefs about the distribution of preferences. This reflects the way that voters process the informational content of the polls. We find clear evidence that our subjects use this information to adjust their beliefs in the right direction. At the same time, they do not use the poll outcome to the extent that is theoretically possible. In short, we observe a strong information effect, but not all information is used effectively. As a mechanism that underlies the coordination effect of polls, we have considered the copy-and-paste strategy where a 'successful' protest in the polls leads to protest voters making the same choice in the election as in the poll. Once again, we find that our participants did not apply this strategy to the full extent possible, but they do apply it on a large scale. In sum, we find evidence that the information effect works through participants' updating of beliefs about the underlying distribution of preferences while the coordination effect works because they apply the copy-and-paste strategy.

The empirical method we use is a laboratory experiment with small, committee size, electorates. This extrapolates from behaviour in real-world elections by abstracting from contextual factors that motivate voting decisions in the field and involves electorate sizes that are much smaller than those used to elect candidates to public office. Experimental control, however, allows us to focus on the strategic calculus that protest voters need to undertake to maximise the chances of a successful protest. The strategic environment that protest voters face in our controlled laboratory environment experiment is similar to that which they face in a large-scale election: such voters must form expectations about the preferences and voting intentions of the other voters in the electorate and subsequently cast their vote to try to maximise the likelihood of a successful protest. We expect the way in which voters' behaviour responds to those expectations on a small scale to be similar to the one they do on a large scale and so our laboratory study enables us to understand the voter behaviour of interest. Of course, future research should investigate the robustness of our findings to elections on a larger scale.

As this paper demonstrates with a number of motivating examples, protest voting can impact electoral outcomes. In this paper, we contribute to an emerging experimental literature that studies the conditions under which protest voting is most prevalent and find that subjects do cast a protest vote with greater frequency when they are able to access the result of an opinion poll before making voting decisions (as in both the *Exogenous* and *Endogenous* treatments). From a practical perspective, our results are relevant to regulators who regulate the publication of opinion polls in the lead-up to elections. Embargoes on opinion polls have been shown to dampen the incidence of strategic voting in elections (Lago et al., 2015) and our results may suggest that such embargoes could also discourage potential protest voters from casting a successful protest vote, due to the increased likelihood of coordination failure and the risk of electoral accident. In addition, the impact of protest voting is important in the context of the UK EU Membership Referendum. In the run-up to the referendum date, most opinion polls indicated a win for the remain side (Duncan, 2016). As such, opinion polls may have encouraged protest voters to underestimate the true underlying level of support for the Leave side by, for example, causing them to incorrectly infer that they were in an environment a scenario akin to our scenario 2 instead of scenario 3 and thereby concluding that protest voting was a safer strategy (in terms of avoiding overshooting) than it actually was. As such, in electoral situations where incentives to engage in protest voting are abound, the availability of opinion poll results may impact on electoral outcomes through the (more effective) occurrence of protest voting.

# Bibliography


Adams, J., Dow, J., & Merrill III, S. (2006). The Political Consequences of Alienation-Based and Indifference-Based Voter Abstention: Applications to Presidential Elections. *Political behavior*, 28(1), 65–86.

Agranov, M., Goeree, J. K., Romero, J., & Yariv, L. (2018). What Makes Voters Turn Out: The Effects of Polls and Beliefs. *Journal of the European Economic Association*, 16(3), 825–856.

Alabrese, E. & Fetzer, T. (2018). *Who is NOT voting for Brexit anymore?* CAGE Online Working Paper Series 394.

Alvarez, R. M., Kiewiet, D. R., & Núñez, L. (2018). A Taxonomy of Protest Voting. *Annual Review of Political Science*, 21(1), 135–154.

Andonie, C. & Kuzmics, C. (2012). Pre-election polls as strategic coordination devices. *Journal of economic behavior & organization*, 84(2), 681–700.

Bergh, J. (2004). Protest Voting in Austria, Denmark, and Norway. *Scandinavian Political Studies*, 27(4), 367–389.

Blais, A. (2014). *To Vote or Not to Vote The Merits and Limits of Rational Choice Theory*. University of Pittsburgh Press.

Boukouras, A., Jennings, W., Li, L., & Maniadis, Z. (2023). Can biased polls distort electoral results? evidence from the lab. *European Journal of Political Economy*, 78, 102383.

Bowler, S. & Lanoue, D. J. (1992). Strategic and Protest Voting for Third Parties: The Case of the Canadian NDP. *The Western Political Quarterly*, 45(2), 485–499.

Brennan, G. & Hamlin, A. (1998). Expressive Voting and Electoral Equilibrium. *Public choice*, 95(1/2), 149–175.

Castanheira, M. (2003). Why Vote For Losers? *Journal of the European Economic Association*, 1(5), 1207–1238.

Charness, G., Garcia, T., Offerman, T., & Villeval, M. (2020). Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory? *Journal of risk and uncertainty*, 60(2), 99–123.

Charness, G. & Gneezy, U. (2010). Portfolio Choice and Risk Attitudes: An Experiment. *Economic Inquiry*, 48(1), 133–146.

Dawes, R. M., Orbell, J. M., Simmons, R. T., & Van De Kragt, A. J. C. (1986). Organizing Groups for Collective Action. *The American political science review*, 80(4), 1171–1185.

Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, 65(2), 135–150.

Duncan, P. (2016). How the pollsters got it wrong on the eu referendum.

Engelen, B. (2006). Solving the Paradox: The Expressive Rationality of the Decision to Vote. *Rationality and Society*, 18(4), 419–441.

Feddersen, T. J. & Pesendorfer, W. (1996). The swing voter's curse. *The American Economic Review*, 86(3), 408–424.

Feddersen, T. J. & Pesendorfer, W. (1999). Abstention in elections with asymmetric information and diverse preferences. *The American Political Science Review*, 93(2), 381–398.

Feltham, O., Schram, A., & Sloof, R. (2021). The Role of Opinion Polls in Coordination Amongst Protest Voters: An Experimental Study (Master's thesis). *University of Amsterdam.*

Fetzer, T. (2019). Did Austerity Cause Brexit? *The American economic review*, 109(11), 3849–3886.

Fey, M. (1997). Stability and Coordination in Duverger's Law: A Formal Model of Preelection Polls and Strategic Voting. *The American Political Science Review*, 91(1), 135–147.

Fisher, S. D. (2004). Definition and Measurement of Tactical Voting: The Role of Rational Choice. *British Journal of Political Science*, 34(1), 152–166.

Forsythe, R., Myerson, R., Rietz, T., & Weber, R. (1993). An experiment on coordination in multi-candidate elections: The importance of polls and election histories. *Social choice and welfare*, 10(3), 223–247.

Forsythe, R., Myerson, R., Rietz, T., & Weber, R. (1996). An Experimental Study of Voting Rules and Polls in Three-Way Elections. *International Journal of Game Theory*, 25, 355–383.

Franklin, M., Niemi, R., & Whitten, G. (1994). The Two Faces of Tactical Voting. *British Journal of Political Science*, 24(4), 549–557.

Frankovic, K., Johnson, T., & Stavrakantonaki, M. (2018). *Freedom to Conduct Opinion Polls: A 2017 Worldwide Update*. ESOMAR and WAPOR.

Gneezy, U. & Potters, J. (1997). An Experiment on Risk Taking and Evaluation Periods. *The Quarterly Journal of Economics*, 112(2), 631–645.

Großer, J. & Schram, A. (2010). Public Opinion Polls, Voter Turnout, and Welfare: An Experimental Study. *American journal of political science*, 54(3), 700–717.

Heath, O. & Goodwin, M. (2017). The 2017 General Election, Brexit and the Return to Two-Party Politics: An Aggregate-Level Analysis of the Result. *The Political Quarterly*, 88(3), 345–358.

Hossain, T. & Okui, R. (2013). The Binarized Scoring Rule. *The Review of Economic Studies*, 80(3), 984–1001.

Kang, W.-T. (2004). Protest Voting and Abstention Under Plurality Rule Elections. *Journal of Theoretical Politics*, 16, 102 – 79.

Kselman, D. & Niou, E. (2011). Protest voting in plurality elections: a theory of voter signaling. *Public Choice*, 148(3), 395–418.

Lago, I., Guinjoan, M., & Bermúda, S. (2015). Regulating Disinformation: Poll Embargo and Electoral Coordination. *The Public Opinion Quarterly*, 79(4), 932–951.

Levine, D. K. & Palfrey, T. R. (2007). The Paradox of Voter Participation? A Laboratory Study. *The American Political Science Review*, 101(1), 143–158.

Louis, P., Troumpounis, O., Tsakas, N., & Xefteris, D. (2020). Protest voting in the laboratory. (288072952).

Louis, P., Troumpounis, O., Tsakas, N., & Xefteris, D. (2022). Coordination with preferences over the coalition size. *Journal of Economic Behavior & Organization*, 194, 105–123.

Myatt, D. P. (2017). A Theory of Protest Voting. *The Economic Journal*, 127(603), 1527–1567.

Offerman, T., Sonnemans, J., & Schram, A. (1996). Value Orientations, Expectations and Voluntary Contributions in Public Goods. *The Economic Journal*, 106(437), 817–845.

Piketty, T. (2000). Voting as Communicating. *The Review of Economic Studies*, 67(1), 169–191.

Razin, R. (2003). Signaling and Election Motivations in a Voting Model with Common Values and Responsive Candidates. *Econometrica*, 71(4), 1083–1119.

Riker, W. H. & Ordeshook, P. C. (1968). A Theory of the Calculus of Voting. *The American Political Science Review*, 62(1), 25–42.

Schimpf, C. H. (2019). Anticipated Election Result and Protest Voting: Why and When Canadian Voters Signal Discontent. *Canadian Journal of Political Science*, 52(4), 847–863.

Schimpf, C. H. (2021). *The Micro Logic of Tactical Protest Voting: A Comparative Analysis*. Wiesbaden: Springer VS.

Simon, H. A. (1954). Bandwagon and Underdog Effects and the Possibility of Election Predictions. *The Public Opinion Quarterly*, 18(3), 245–253.

Sugden, R. (1986). *The economics of rights, co-operation and welfare*. Oxford: Basil Blackwell.

Tyszler, M. & Schram, A. (2016). Information and strategic voting. *Experimental economics : a journal of the Economic Science Association*, 19(2), 360–381.

van de Kragt, A. J. C., Orbell, J. M., & Dawes, R. M. (1983). The Minimal Contributing Set as a Solution to Public Goods Problems. *The American political science review*, 77(1), 112–122.

Wang, B. & Zhou, Z. (2023). Informational feedback between voting and speculative trading. *Games and Economic Behavior*, 138, 387–406.

Wilson, A. & Vespa, E. (2018). *Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language*. Technical report, Working paper.

# Appendices

*To be published online*

# A Theoretical equilibrium analysis

## A.1 Full information about the type composition of the electorate

Let the composition of the electorate $\mu = (m, a, p)$ be common knowledge, with $m + a + p = n$. The critical range (of votes for $A$) where the protest succeeds equals $[\underline{t}, \bar{t}]$, with $\underline{t} < \bar{t} = \frac{n-1}{2}$. If all other protest voters individually cast a protest vote with probability $\pi$, the net benefit for voter $i$ of doing so as well is given by equation (1) in the main text. In this expression $t_{-i}$ denotes the overall number of actual votes for $A$ among all other $n-1$ voters. Based on the binomial distribution the two pivotal probabilities in (1) can be characterised and the following proposition is obtained.

**Proposition** 1. Suppose $\mu = (m, a, p)$ is common knowledge. Based on $\mu$ one of four different scenarios applies:

**(i)** *Indifferent scenario*. If either (i) $a > \bar{t}$, or (ii) $\underline{t} \leq a \leq a + p \leq \bar{t}$, or (iii) $a + p < \underline{t}$, protest voters are never pivotal and any equilibrium $\pi^e \in [0, 1]$ can occur;

**(ii)** *Protest scenario*. If $a < \underline{t} \leq a + p \leq \bar{t}$, there is a unique symmetric Nash equilibrium in weakly undominated strategies in which $\pi^e = 1$;

**(iii)** *Support scenario*. If $\underline{t} \leq a \leq \bar{t} < a + p$, there is a unique symmetric Nash equilibrium in weakly undominated strategies in which $\pi^e = 0$;

**(iv)** *Mixed scenario*. If $a < \underline{t} < \bar{t} < a + p$, there is a unique symmetric Nash equilibrium in weakly undominated strategies satisfying responsiveness (cf. Louis et al, 2022) in which:

$$\pi^e = \frac{1}{1 + \left[ \frac{(\underline{t}-1-a)! \cdot (p+a-\underline{t})!}{(\bar{t}-a)! \cdot (p+a-1-\bar{t})!} \cdot \frac{1+s}{s} \right]^{\frac{1}{\bar{t}+1-\underline{t}}}} \tag{A.1}$$

**Proof**. We use $\underline{\tau}(\pi) \equiv \Pr(t_{-i} = \underline{t} - 1 \,|\, \pi, \mu)$ and $\overline{\tau}(\pi) \equiv \Pr(t_{-i} = \bar{t} \,|\, \pi, \mu)$ as shorthand notation for the two pivotal probabilities. Equation (1) then becomes $NB_i(\text{protest vote} \,|\, \pi, \mu) = \underline{\tau}(\pi) \cdot s - \overline{\tau}(\pi) \cdot (1 + s)$. In the first, indifferent scenario we have $\underline{\tau}(\pi) = \overline{\tau}(\pi) = 0$ for all $\pi \in [0, 1]$, because either (i) $A$ always wins the election irrespective of how protest voters vote ($a > \bar{t}$), (ii) $M$ necessarily wins with the protest necessarily succeeding ($\underline{t} \leq a \leq a + p \leq \bar{t}$), or (iii) $M$

always wins without the protest ever succeeding because A's potential vote base is too small ($a+p<\underline{t}$). In the second protest scenario it necessarily holds that $\overline{\tau}(\pi)=0$ for all $\pi\in[0,1]$, because $a+p\leq\overline{t}$ implies that candidate $M$ necessarily wins the election. At the same time, $a<\underline{t}\leq a+p$ implies that the overall number of protest votes from protest voters matters for whether the protest succeeds or not (i.e. $\underline{\tau}(\pi)>0$ for all $\pi\in(0,1)$). In that case voting for $A$ is weakly dominant and hence $\pi^e=1$. Similarly, if $\underline{t}\leq a\leq\overline{t}<a+p$ it holds that $\underline{\tau}(\pi)=0$ for all $\pi\in[0,1]$, but the election outcome is not determined yet by the number of $a$ and $m$ voters. In that case voting for $M$ is weakly dominant and $\pi^e=0$ is the unique symmetric NE in weakly undominated strategies.

Finally, suppose $a<\underline{t}<\overline{t}<a+p$. From the binomial distribution we immediately obtain:

$$\underline{\tau}(\pi)\equiv\Pr(t_{-i}=\underline{t}-1\,|\,\pi,\mu)=\binom{p-1}{\underline{t}-1-a}\cdot\pi^{(\underline{t}-1-a)}\cdot(1-\pi)^{(p-\underline{t}+a)}$$

and

$$\overline{\tau}(\pi)\equiv\Pr(t_{-i}=\overline{t}\,|\,\pi,\mu)=\binom{p-1}{\overline{t}-a}\cdot\pi^{(\overline{t}-a)}\cdot(1-\pi)^{(p-1-\overline{t}+a)}$$

From these we also have:

$$\frac{\underline{\tau}(\pi)}{\overline{\tau}(\pi)}=\frac{(\overline{t}-a)!\cdot(p-1-\overline{t}+a)!}{(\underline{t}-1-a)!\cdot(p-\underline{t}+a)!}\cdot\left(\frac{1-\pi}{\pi}\right)^{\overline{t}+1-\underline{t}}$$

First consider mixed equilibria $\pi^e\in(0,1)$. Such an interior equilibrium is characterised by: $\underline{\tau}(\pi^e)\cdot s-\overline{\tau}(\pi^e)\cdot(1+s)=0$, i.e. $\frac{\underline{\tau}(\pi^e)}{\overline{\tau}(\pi^e)}=\frac{1+s}{s}$. As $\frac{\underline{\tau}(\pi)}{\overline{\tau}(\pi)}$ is strictly decreasing in $\pi$, positive for small values of $\pi$ close to zero and negative for large values of $\pi$ close to one, a unique root to this equation exists. Solving for this root yields (A.1).

Next, suppose $\pi=0$. In that case $\overline{\tau}(0)=0$ necessarily. If $\underline{t}-1-a>0$, then $\underline{\tau}(0)=0$ as well, making voting for $M$ a (weak) best response.[1] Hence, if $a<\underline{t}-1$ another pure equilibrium exists side-by-side the mixed one in which $\pi^e=0$. This pure equilibrium is not *responsive* in the sense of Louis et al (2022); protest voters employ the exact same strategy as mainstream voters do. Finally, suppose $\pi=1$. In that case $\underline{\tau}(0)=0$ necessarily. If $p-1-\overline{t}+a>0$, then $\overline{\tau}(0)=0$ as well, making voting for $A$ a (weak) best response. Hence, for $p+a>\overline{t}+1$ another equilibrium exists side-by-side the mixed one in which $\pi^e=1$. Similar to Louis et al (2022), if both responsive ($\pi^e\in(0,1)$) and non-responsive ($\pi^e=0$ or $\pi^e=1$) equilibria exist side by side, we focus on the former. QED

[1] Note that for $\underline{t}-1-a=0$ term $\pi^{(\underline{t}-1-a)}$ equals $0^0=1$, while for $\underline{t}-1-a>0$ this term becomes $0^{(\underline{t}-1-a)}=0$.

If $\underline{t}-1-a=p+a-1-\bar{t}$, i.e. if $\underline{t}-a=\bar{t}+1-m$, expression (A.1) reduces to

$$\pi^e = \frac{1}{1+\left[\frac{1+s}{s}\right]^{\frac{1}{\bar{t}+1-\underline{t}}}} \tag{A.2}$$

In that case necessarily $\pi^e < \frac{1}{2}$. Intuitively, when $\underline{t}-a=\bar{t}+1-m$, the minimum number of protest voters needed to secure the protest's success (by voting for $A$) equals the minimum number of protest voters needed to secures $M$'s victory (by voting for $M$). In that case, if all other protest voters would vote for $A$ and $M$ with equal probabilities ($\pi = \frac{1}{2}$), the probability for voter $i$ being pivotal for the protest's success equals the one for being pivotal for the electoral outcome. Roughly put, the electoral composition is 'symmetric', such that in principle protest voters are equally likely to be pivotal either way. Given that the disutility of overshooting the critical range $(1+s)$ is larger than the disutility of undershooting it $(s)$, voter $i$ then prefers to vote for $M$. This explains why $\pi^e < \frac{1}{2}$ in that case. Our experimental parameters are such that in our mixed scenario indeed $\underline{t}-a=\bar{t}+1-m\ (=1)$. With $\bar{t}+1-\underline{t}=2$ equation (A.2) then reduces to (2) in the main text.

## A.2 No information about type composition

In the remainder of this appendix we focus on the specific electoral setting used in the experiment. That is, let $n=9$, $\underline{t}=3$ and $\bar{t}=4$. Moreover, three different electoral compositions are possible:

Table A1: Possible scenarios and their frequency

| Scenario | Motive | $m$ | $a$ | $p$ | Frequency |
|---|---|---|---|---|---|
| 1 | protest | 6 | 0 | 3 | $\rho_1$ |
| 2 | mixed | 4 | 2 | 3 | $\rho_2$ |
| 3 | support | 2 | 4 | 3 | $\rho_3$ |

The probability of these three scenarios occurring are $\rho_1$, $\rho_2$ and $\rho_3$, respectively. As before, let $\pi$ denote the probability with which an individual protest voter votes for $A$. Assuming all other protest voters individually cast a protest vote with probability $\pi$, the net benefit for protest voter $i$ of casting a protest vote equals (cf. equation (3) in the main text):

$$\begin{aligned} NB_i\,(\text{protest vote}\,|\,\pi,\rho) &= \rho_1 \cdot \left\{\pi^2 \cdot s\right\} + \rho_2 \cdot \left\{(1-\pi)^2 \cdot s - \pi^2 \cdot (1+s)\right\} \\ &\quad + \rho_3 \cdot \left\{-(1-\pi)^2 \cdot (1+s)\right\} \\ &= \pi^2 \cdot [\rho_1 \cdot s - \rho_2 \cdot (1+s)] + (1-\pi)^2 \cdot [\rho_2 \cdot s - \rho_3 \cdot (1+s)] \end{aligned}$$

Based on this expression, we obtain the following proposition, which gives the (unique) symmetric equilibrium if a condition on the relative likelihood of the three different electoral scenarios is met.

**Proposition 2**. Suppose $\rho_2 \geq \min\left\{\frac{s}{1+s}\cdot\rho_1, \frac{1+s}{s}\cdot\rho_3\right\}$. If $\rho_2 = \frac{s}{1+s}\cdot\rho_1 \leq \frac{1+s}{s}\cdot\rho_3$, then $\pi^e = 0$ while for $\rho_2 = \frac{1+s}{s}\cdot\rho_3 \leq \frac{s}{1+s}\cdot\rho_1$ we have $\pi^e = 1$. If $\rho_2 > \min\left\{\frac{s}{1+s}\cdot\rho_1, \frac{1+s}{s}\cdot\rho_3\right\}$, then:

$$\pi^e = \frac{\sqrt{\max\left\{\rho_2 s - \rho_3(1+s), 0\right\}}}{\sqrt{\max\left\{\rho_2 s - \rho_3(1+s), 0\right\}} + \sqrt{\max\left\{\rho_2(1+s) - \rho_1 s, 0\right\}}} \tag{A.3}$$

**Proof**. Let $A \equiv \rho_1 \cdot s - \rho_2 \cdot (1+s)$ and $B \equiv \rho_2 \cdot s - \rho_3 \cdot (1+s)$. We then have:

$$NB_i\left(\text{protest vote}\,|\,\pi, \rho\right) = A\cdot\pi^2 + B\cdot(1-\pi)^2$$

Aside from knife-edge cases (discussed below), four different cases can be distinguished based on the signs of $A$ and $B$. If $A > 0$ and $B > 0$, then $NB_i\left(\text{protest vote}\,|\,\pi,\rho\right) > 0$ for all $\pi$, and hence necessarily $\pi^e = 1$. Similarly, if $A < 0$ and $B < 0$, then $NB_i\left(\text{protest vote}\,|\,\pi,\rho\right) < 0$ for all $\pi$, and hence necessarily $\pi^e = 0$. In case $A < 0$ and $B > 0$, $NB_i\left(\text{protest vote}\,|\,\pi,\rho\right)$ is strictly decreasing in $\pi$ for all $\pi \in [0,1]$, positive for $\pi = 0$ and negative for $\pi = 1$. Hence necessarily $\pi^e \in (0,1)$. This interior level of $\pi^e$ follows from $NB_i\left(\text{protest vote}\,|\,\pi,\rho\right) = 0$. This requires $\left[\frac{\pi}{1-\pi}\right]^2 = -\frac{B}{A} = \frac{\rho_2 s - \rho_3(1+s)}{\rho_2(1+s)-\rho_1 s}$. Rewriting gives $\pi^e = \frac{\sqrt{\rho_2 s - \rho_3(1+s)}}{\sqrt{\rho_2 s-\rho_3(1+s)}+\sqrt{\rho_2(1+s)-\rho_1 s}}$. Note that these first three cases are all fully captured by expression (A.3). The fourth case $A > 0$ and $B < 0$ applies when $\rho_1 > \rho_2 \cdot \frac{1+s}{s}$ and $\rho_3 > \rho_2 \cdot \frac{s}{1+s}$. It therefore does not occur under the supposition $\rho_2 \geq \min\left\{\frac{s}{1+s}\cdot\rho_1, \frac{1+s}{s}\cdot\rho_3\right\}$ made.[2]

We finally consider the remaining knife-edge cases where either $A = 0$ or $B = 0$, i.e. where the supposition made holds with an equality: $\rho_2 = \min\left\{\frac{s}{1+s}\cdot\rho_1, \frac{1+s}{s}\cdot\rho_3\right\}$. The case $\rho_2 = \frac{s}{1+s}\cdot\rho_1 \leq \frac{1+s}{s}\cdot\rho_3$ corresponds to $A = 0$ and $B \leq 0$. The net benefit of a protest vote then equals $B\cdot(1-\pi)^2 \leq 0$, implying $\pi^e = 0$. Similarly, for $\rho_2 = \frac{1+s}{s}\cdot\rho_3 \leq \frac{s}{1+s}\cdot\rho_1$ we have $B = 0$ and $A \geq 0$. The net benefit of a protest vote then equals $A\cdot\pi^2 \geq 0$, implying $\pi^e = 1$.[3] QED

For our choice of prior probabilities $(\rho_1, \rho_2, \rho_3) = (0.4, 0.5, 0.1)$ and of protest success bonus $s = 1$ in the experiment, we have that both $\rho_2 > \frac{s}{1+s}\cdot\rho_1$ and $\rho_2 > \frac{1+s}{s}\cdot\rho_3$ hold. Proposition 2 thus applies for the control treatment without polls and, given that for these parameter choices in fact $\rho_2 \geq \max\left\{\frac{s}{1+s}\cdot\rho_1, \frac{1+s}{s}\cdot\rho_3\right\}$, (A.3) reduces to (4) in the main text. The proposition is also of direct use for the treatment with exogenous polls, just replacing $\rho_j$ with $\widehat{\rho}_j(d;\sigma)$

[2] Just for completeness, in this case three equilibria exist side by side, with either $\pi^e = 0$, $\pi^e = 1$, and an interior one satisfying $\pi^e = \frac{\sqrt{\rho_3(1+s)-\rho_2 s}}{\sqrt{\rho_3(1+s)-\rho_2 s}+\sqrt{\rho_1 s-\rho_2(1+s)}}$. This follows because for $A > 0$ and $B < 0$, $NB_i\left(\text{protest vote}\,|\,\pi,\rho\right)$ is strictly increasing in $\pi$ for all $\pi \in [0,1]$, negative for $\pi = 0$ and positive for $\pi = 1$.

[3] Clearly, if both $A = 0$ and $B = 0$, the net benefit of a protest vote always equals zero and any $\pi^e \in [0,1]$ goes. This may only happen for non-generic parameter combinations where $\rho_2 = \frac{1+s}{s}\cdot\rho_3 = \frac{s}{1+s}\cdot\rho_1$ together with (per $\rho_1+\rho_2+\rho_3 = 1$) $\rho_1 = \frac{(1+s)^2}{1+3s+3s^2}$. Such extreme knife edge cases (effectively equivalent to the indifferent scenario in Proposition 1) do not occur in our experiment.

for all $j$. Also in this case the condition on (now) $\widehat{\rho}_2(d;\sigma)$ is satisfied for all $(d,\sigma)$, because after the polls always one of the more "extreme" scenarios can be excluded based on observing $d$. That is, $\widehat{\rho}_3(d;\sigma)=0$ for $d\leq 3$ and $\widehat{\rho}_1(d;\sigma)=0$ for $d\geq 4$.

Finally, if $\rho_1=\left[\frac{1+s}{s}\right]^2\rho_3$, equilibrium probability $\pi^e$ as characterised in (A.3) corresponds to the case where you know that scenario 2 applies for sure, i.e. to $\pi^e=\frac{\sqrt{s}}{\sqrt{s}+\sqrt{1+s}}$. This condition is met by our choice of prior probabilities $(\rho_1,\rho_2,\rho_3)=(0.4,0.5,0.1)$ and $s=1$ in the experiment. For these parameters we obtain $\pi^e=\frac{1}{1+\sqrt{2}}\simeq 0.41$ under both No Information and the case where you know that scenario 2 applies for sure ($\rho_2=1$).

## A.3 Exogenous opinion polls

In this subsection we analyse the situation in which voters observe the outcome of a pre-election poll before they cast their vote. We consider the case of exogenous polls where voters observe the result from a poll in which they did not take part themselves. That is, they observe the poll result from an electorate that is in the exact same situation as they are, except for the fact that the other electorate does take part in an endogenous poll. In order to gauge the informational value of the polls outcome received, voters should form beliefs about how this other electorate behaved in the polls. Let $\sigma\in[0,1]$ denote the (exogenous) probability with which an individual protest voter in the other electorate declares an intention to cast a protest vote at the polls. Posterior beliefs $\widehat{\rho}_j(d;\sigma)$ after observing overall $d$ declarations to vote for $A$ in the other electorate then follow from applying Bayes' rule.

The table below provides these posterior beliefs after each possible poll result. In turn inserting these in Proposition 2 above, we immediately obtain the equilibrium probabilities $\pi^e(d;\sigma)$ with which an individual protest voter casts a protest vote in the election (see the final column). The point predictions for our parameter choices reported in the main text follow from this table, taking $\sigma=0.489$. The latter specific value for $\sigma$ follows from our analysis for the endogenous polls treatment, to which we turn next.

## A.4 Endogenous opinion polls

Let $\sigma$ now denote the endogenous probability with which an individual protest voter declares an intention to cast a protest vote at the polls. We proceed by backwards induction, by first analysing the equilibrium in the continuation game after the outcome of the polls is made public and protest voters simultaneously have to cast their actual vote. The equilibria of this continuation game depend on the posterior beliefs after the polls. In general, posterior beliefs now not only depend on the overall polling outcome $d$, but also on protest voter $i$'s own polling behaviour $d_i\in\{0,1\}$: i.e., we get $\widehat{\rho}_j(d,d_i;\sigma)$. Note that by definition it must hold that $\widehat{\rho}_j(d,1;\sigma)=\widehat{\rho}_j(d-1,0;\sigma)$. Intuitively, when making inferences about which scenario applies

Table A2: Posterior beliefs and equilibrium protest voting with Exogenous polls

| $d$ | $\hat{\rho}_1(d;\sigma)$ | $\hat{\rho}_2(d;\sigma)$ | $\hat{\rho}_3(d;\sigma)$ | $\pi^e(d;\sigma)$ |
|---|---|---|---|---|
| $0,1$ | 1 | | | 1 |
| 2 | $\frac{\rho_1\cdot 3\sigma^2(1-\sigma)}{\rho_1\cdot 3\sigma^2(1-\sigma)+\rho_2\cdot(1-\sigma)^3}$ | $\frac{\rho_2\cdot(1-\sigma)^3}{\rho_1\cdot 3\sigma^2(1-\sigma)+\rho_2\cdot(1-\sigma)^3}$ | | $\frac{\sqrt{\hat{\rho}_2 s}}{\sqrt{\hat{\rho}_2 s}+\sqrt{\max\{\hat{\rho}_2(1+s)-\hat{\rho}_1 s,0\}}}$ |
| 3 | $\frac{\rho_1\cdot\sigma^3}{\rho_1\cdot\sigma^3+\rho_2\cdot 3\sigma(1-\sigma)^2}$ | $\frac{\rho_2\cdot 3\sigma(1-\sigma)^2}{\rho_1\cdot\sigma^3+\rho_2\cdot 3\sigma(1-\sigma)^2}$ | | $\frac{\sqrt{\hat{\rho}_2 s}}{\sqrt{\hat{\rho}_2 s}+\sqrt{\max\{\hat{\rho}_2(1+s)-\hat{\rho}_1 s,0\}}}$ |
| 4 | | $\frac{\rho_2\cdot 3\sigma^2(1-\sigma)}{\rho_2\cdot 3\sigma^2(1-\sigma)+\rho_3\cdot(1-\sigma)^3}$ | $\frac{\rho_3\cdot(1-\sigma)^3}{\rho_2\cdot 3\sigma^2(1-\sigma)+\rho_3\cdot(1-\sigma)^3}$ | $\frac{\sqrt{\max\{\hat{\rho}_2 s-\hat{\rho}_3(1+s),0\}}}{\sqrt{\max\{\hat{\rho}_2 s-\hat{\rho}_3(1+s),0\}}+\sqrt{\hat{\rho}_2(1+s)}}$ |
| 5 | | $\frac{\rho_2\cdot\sigma^3}{\rho_2\cdot\sigma^3+\rho_3\cdot 3\sigma(1-\sigma)^2}$ | $\frac{\rho_3\cdot 3\sigma(1-\sigma)^2}{\rho_2\cdot\sigma^3+\rho_3\cdot 3\sigma(1-\sigma)^2}$ | $\frac{\sqrt{\max\{\hat{\rho}_2 s-\hat{\rho}_3(1+s),0\}}}{\sqrt{\max\{\hat{\rho}_2 s-\hat{\rho}_3(1+s),0\}}+\sqrt{\hat{\rho}_2(1+s)}}$ |
| $6,7$ | | | 1 | 0 |

based on the overall polling outcome, an individual voter corrects for their own polling statement. Inferences of individual voter $i$ are thus based on the 'net' poll result $d-d_i$. Applying Bayes' rule yields posterior beliefs $\overline{\rho}_j(d-d_i;\sigma)$ based on these net polls results, which are reported in the table below (where $\overline{\rho}$ refers to beliefs based on net poll results). Note that it holds that $\widehat{\rho}_j(d,d_i;\sigma)=\overline{\rho}_j(d-d_i;\sigma)$ for $j=1,2,3$.

Given that an overall poll outcome equal to $d$ implies a different 'net' poll outcome for a voter who declared to vote for $M$ in the polls ($d_i=0$) than for someone who declared a vote for $A$ ($d_i=1$), voters may well hold different posterior beliefs (depending on the value of $d$). Fully rational voters will realise this and the equilibrium analysis takes this into account.

### A.4.2 Equilibrium voting behaviour after the polls

The voting continuation game in general allows for multiple equilibria. As discussed in the main text, we focus on the more 'reasonable' equilibria by making two equilibrium refinement assumptions:

**(i)** if protest voters have a weakly dominant strategy, they will behave accordingly;

**(ii)** if the polls perfectly 'solve' the coordination problem, all protest voters simply vote as they did in the polls.

Table A3: Posterior beliefs with Endogenous polls

| $d-d_i$ | $\overline{\rho}_1(d-d_i;\sigma)$ | $\overline{\rho}_2(d-d_i;\sigma)$ | $\overline{\rho}_3(d-d_i;\sigma)$ |
|---|---|---|---|
| 0,1 | 1 | | |
| 2 | $\frac{\rho_1\sigma^2}{\rho_1\sigma^2+\rho_2(1-\sigma)^2}$ | $\frac{\rho_2(1-\sigma)^2}{\rho_1\sigma^2+\rho_2(1-\sigma)^2}$ | |
| 3 | | 1 | |
| 4 | | $\frac{\rho_2\sigma^2}{\rho_2\sigma^2+\rho_3(1-\sigma)^2}$ | $\frac{\rho_3(1-\sigma)^2}{\rho_2\sigma^2+\rho_3(1-\sigma)^2}$ |
| 5,6,7 | | | 1 |

Note: it holds that $\widehat{\rho}_j(d,d_i;\sigma)=\overline{\rho}_j(d-d_i;\sigma)$ for $j=1,2,3$.

Assumption (i) fully characterises equilibrium behaviour when either $d\in\{0,1\}$ or when $d\in\{6,7\}$. For these poll outcomes all protest voters share the same beliefs and they either unanimously conclude that $\hat{\rho}_1(d,d_i;\sigma)=1$ or that $\hat{\rho}_3(d,d_i;\sigma)=1$, respectively. In the former case $\pi^e(d,d_i;\sigma)=1$ is weakly dominant, in the latter case $\pi^e(d,d_i;\sigma)=0$. Assumption (ii) fully characterises equilibrium behaviour when $d=3$ or $d=4$; in these instances, if a protest voter indicated $d_i=1$ in the polls, she will choose $\pi^e(d,d_i;\sigma)=1$, otherwise $\pi^e(d,d_i;\sigma)=0$. Finally, for $d=2$ a protest voter who declared $d_i=1$ will infer that $\hat{\rho}_1(2,1;\sigma)=1$ and, per Assumption (i), choose $\pi^e(2,1;\sigma)=1$. Similarly, a protest voter who declared $d_i=0$ while $d=5$ overall, will infer that $\hat{\rho}_3(5,0;\sigma)=1$ and choose $\pi^e(5,0;\sigma)=0$. Overall, only $\pi^e(2,0;\sigma)$ and $\pi^e(5,1;\sigma)$ thus remain to be characterised. The following lemma does so.

**Lemma** 1. It holds that $\pi^e(2,0;\sigma)=\min\left\{-s+\sqrt{s^2+s+\left[\frac{\rho_1\sigma^2}{\rho_2(1-\sigma)^2}\right]\cdot s},\,1\right\}$ and $\pi^e(5,1;\sigma)=\max\left\{-s+\sqrt{s^2+s-\left[\frac{\rho_3(1-\sigma)^2}{\rho_2\sigma^2}\right]\cdot(1+s)},\,0\right\}$.

**Proof**. First consider $d=2$. For a protest voter with $d_i=1$ we have $\pi^e(2,1;\sigma)=1$, as explained above the lemma. A protest voter with $d_i=0$ believes that there are two possibilities: (A) $a=0$ and the two other protest voters both polled $d_{-i}=1$, or (B) $a=2$ and the two other protest voters both polled $d_{-i}=0$. In case (A) the two other voters then will choose their weakly dominant strategy $\pi^e(2,1;\sigma)=1$ in the voting stage, while in case (B) both other protest voters will use $\pi(2,0;\sigma)$ (we thus assume symmetry). The net payoffs of voting for $A$ for the protest voter under consideration then equals:

$$NB_i\big(\text{protest vote}\,|\,\pi,\hat{\rho}\left(\sigma\right),d=2,d_i=0\big)=$$
$$\hat{\rho}_1\left(2,0;\sigma\right)\cdot s+\hat{\rho}_2\left(2,0;\sigma\right)\cdot\left[\left(1-\pi\left(2,0;\sigma\right)\right)^2\cdot s\pi\left(2,0;\sigma\right)^2\cdot\left(1+s\right)\right].$$

The equilibrium value of $\pi\left(2,0;\sigma\right)$ follows from setting the above net benefits to zero. Together with $\hat{\rho}_2\left(2,0;\sigma\right)=\frac{\rho_2(1-\sigma)^2}{\rho_1\sigma^2+\rho_2(1-\sigma)^2}=1-\hat{\rho}_1\left(2,0;\sigma\right)$ this gives the expression in the lemma. (Note that if $\hat{\rho}_2\left(2,0;\sigma\right)$ gets small, no interior solution exists and $\pi^e\left(2,0;\sigma\right)=1$.)

Next consider $d=5$. For a protest voter with $d_i=0$ we have $\pi^e\left(5,0;\sigma\right)=0$ immediately, as $\hat{\rho}_3\left(5,0;\sigma\right)=1$ and voting for $M$ is weakly dominant. A protest voter with $d_i=1$ believes that there are two possibilities: (C) $a=2$ and the two other protest voters polled $d_{-i}=1$ as well, or (D) $a=4$ and the two other protest voters polled $d_{-i}=0$. In case C the other two protest voters will employ $\pi\left(5,1;\sigma\right)$ (symmetry), while in case D the other two protest voters vote mainstream. The net payoff of voting anti-mainstream then equals:

$$NB_i\big(\text{protest vote}\,|\,\pi,\hat{\rho}\left(\sigma\right),d=5,d_i=1\big)=$$
$$\hat{\rho}_2\left(5,1;\sigma\right)\cdot\left[\left(1-\pi\left(5,1;\sigma\right)\right)^2\cdot s-\pi\left(5,1;\sigma\right)^2\cdot\left(1+s\right)\right]-\hat{\rho}_3\left(5,1;\sigma\right)\cdot\left(1+s\right)$$

The equilibrium value of $\pi\left(5,1;\sigma\right)$ follows from setting the above net benefits to zero. Together with $\hat{\rho}_2\left(5,1;\sigma\right)=\frac{\rho_2\sigma^2}{\rho_2\sigma^2+\rho_3(1-\sigma)^2}=1-\hat{\rho}_3\left(5,1;\sigma\right)$ this gives the expression in the lemma. (Note that if $\hat{\rho}_2\left(5,1;\sigma\right)$ gets small, no interior solution exists and $\pi^e\left(5,1;\sigma\right)=0$.) QED.

From the above lemma we have that $\pi^e\left(5,1;\sigma\right)\leq\pi^e\left(2,0;\sigma\right)$ necessarily. This is intuitive, because at $d=5$ the possibility of overshooting at the actual vote (just like in the polls) is larger than for $d=2$. Overall equilibrium voting behaviour after the polls as captured by $\pi^e\left(d,d_i;0\right)$ is summarised in the table below.

### A.4.3 Equilibrium behaviour at the polls

Finally, we consider equilibrium behaviour at the polls. Use $\pi_2=\pi^e\left(2,0;\sigma\right)$ and $\pi_5=\pi^e\left(5,1;\sigma\right)$ as short hand notation. Based on the earlier assumptions (i) and (ii) in the previous subsection, we then obtain:

$$\begin{aligned}NB_i\big(\text{poll for }A\,|\,\pi,\rho\big)=&\;\rho_1\cdot\left\{\left(1-\sigma\right)^2\cdot 0-2\sigma\left(1-\sigma\right)\cdot\left(1-\pi_2\right)\cdot s+\sigma^2\cdot\left(1-\pi_{2i}\right)\cdot s\right\}\\&+\rho_2\cdot\left\{\left(1-\sigma\right)^2\cdot\left(\left[1-\pi_{2i}\right]\cdot\left(1-\pi_2\right)^2\cdot s+\pi_{2i}\cdot\pi_2^2\cdot\left(1+s\right)\right)\right.\\&\left.+2\sigma\left(1-\sigma\right)\cdot 0-\sigma^2\cdot\left(\left[1-\pi_{5i}\right]\cdot\left(1-\pi_5\right)^2\cdot s+\pi_{5i}\cdot\pi_5^2\cdot\left(1+s\right)\right)\right\}\\&+\rho_3\cdot\left\{\sigma^2\cdot 0+2\sigma\left(1-\sigma\right)\cdot\pi_5\cdot\left(1+s\right)-\left(1-\sigma\right)^2\cdot\pi_{5i}\cdot\left(1+s\right)\right\}\end{aligned}$$

Table A4: Equilibrium protest voting with Endogenous polls

| $d$ | $\pi^e(d,0;\sigma)$ | $\pi^e(d,1;\sigma)$ |
|---|---|---|
| 0 | 1 | n.p. |
| 1 | 1 | 1 |
| 2 | $\min\left\{-s+\sqrt{s^2+s+\left[\frac{\rho_1\sigma^2}{\rho_2(1-\sigma)^2}\right]\cdot s},\,1\right\}$ | 1 |
| 3 | 0 | 1 |
| 4 | 0 | 1 |
| 5 | 0 | $\max\left\{-s+\sqrt{s^2+s-\left[\frac{\rho_3(1-\sigma)^2}{\rho_2\sigma^2}\right]\cdot(1+s)},\,0\right\}$ |
| 6 | 0 | 0 |
| 7 | n.p. | 0 |

This expression can be understood as follows. Suppose that scenario 1 applies, which happens with probability $\rho_1$. In that case, if both other protest voters poll mainstream ($d_{-i}=0$), then either $d=0$ (if individual $i$ polls mainstream as well, i.e. if $d_i=0$) or $d=1$ (when $d_i=1$). Either way, $\pi=1$ in the subsequent voting stage for all protest voters, irrespective of what voter $i$ does in the polls. This explains the term $(1-\sigma)^2\cdot 0$. In case only one of the two other protest voters polls $M$ – which happens with probability $2\sigma(1-\sigma)$ – the own polling statement of individual $i$ does matter for subsequent voting behaviour. If $i$ polls $M$, then $d=1$ and in the subsequent voting stage $\pi=1$ for all protest voters, yielding a payoff of $1+s$. If instead $i$ polls $A$, then $d=2$ and both he and the other protest voter polling for $A$ choose $\pi=1$ (see the table in the previous subsection). The other protest voter polling for $M$ chooses to protest vote with probability $\pi_2=\pi^e(2,0;\sigma)$. This gives individual $i$ an expected payoff of $1+\pi_2\cdot s$. The net benefit of polling for $A$ thus equals $(1+\pi_2\cdot s)-(1+s)=-(1-\pi_2)s$. This explains the second term on the first row. Finally, if both other protest voters poll $A$, choosing $d_i=1$ as well leads to $d=3$ and all protest voters choosing $\pi=1$. This yields $1+s$ in payoffs, since the protest succeeds. Choosing $d_i=0$ instead gives $d=2$ and the other two protest voters choosing $\pi=1$. Voter $i$ herself then uses $\pi_2$ in the election stage. This gives expected payoffs $1+\pi_{2i}\cdot s$; here we have used subscript $i$ in the mixing probability $\pi_{2i}$ to indicate that it belongs to $i$ herself. The net benefit of polling $A$ then equals $(1-\pi_{2i})\cdot s$, providing the third term on the first row. All other terms – corresponding to the case where either scenario 2 or 3 applies – follow similarly.

The expression for $NB_i(\text{poll for } A\,|\,\pi,\rho)$ can be simplified by setting $\pi_{2i}=1$ and $\pi_{5i}=0$; this follows because the expected payoffs of a mixed strategy under which you are indifferent, can be calculated by assuming that you choose one of the options between which you mix for sure. (This also explains why in the above expression we separated the mixing probabilities of the other two protest voters from the one of voter $i$.) The polling statement of voter $i$ then only matters to the extent that it changes other protest voter's behaviour. Using this, the above expression reduces to:

$$
\begin{aligned}
NB_i\left(\text{poll for } A \,|\, \pi,\rho\right) = {} & \rho_1 \cdot \{-2\sigma\left(1-\sigma\right)\cdot\left(1-\pi_2\right)\cdot s\} \\
& + \rho_2 \cdot \left\{\left(1-\sigma\right)^2 \cdot \pi_2^2 \cdot \left(1+s\right) - \sigma^2 \cdot \left(1-\pi_5\right)^2 \cdot s\right\} \\
& + \rho_3 \cdot \{2\sigma\left(1-\sigma\right)\cdot \pi_5 \cdot \left(1+s\right)\}
\end{aligned}
$$

Note that this expression is negative for $\sigma = 0$ and positive for $\sigma = 1$, so by the intermediate value theorem there must be at least one root and thus a mixed equilibrium $\sigma^e \in (0,1)$. Since $\pi_2$ and $\pi_5$ themselves depend on $\sigma$ (see Lemma 1), however, it is not a simple quadratic expression in $\sigma$. We thus solve it numerically and find for our parameter choices that $\sigma^e = 0.489$.

# B Experiment instructions: Endogenous treatment

## B.1 Welcome

Thank you for participating in this experiment. In this experiment, you can earn money. The amount of money you earn depends on the decisions you and the other participants make. We will ensure that your final earnings remain confidential; we will not inform other participants of your final earnings.

The session will consist of five parts:

1. Instructions
2. Practice rounds
3. Main Experiment
4. Investment Task
5. Survey

You are now asked to carefully read the instructions. While going through the instructions, you are asked to answer some questions to check your understanding. You will need to answer all of these questions correctly to proceed with the experiment and get paid. If you get stuck after a few attempts, please raise your hand and the experimenter will provide you with assistance. You will also be provided with a handout which summarises the main aspects of the instructions. If you require any assistance during this experiment, please raise your hand and the experimenter will come to assist you.

## B.2 Part 1: Setup

The experiment consists of a series of elections. There are **20 election rounds**, preceded by **5 practice rounds**.

In each election, there are **9 voters** of which **3 (including yourself)** are participants and the decisions of the remaining **6 are programmed automatically**. There are two candidates in the election: A and B. In each election, you will be required to vote for either the A- or the B-candidate. There is no option to abstain, so you must cast a vote for one of these two candidates in order to proceed. The candidate who obtains the highest number of votes wins the election.

The programmed voters will be randomly assigned to be either an A- or a B-supporter. **These programmed voters will be programmed to vote for the candidate that they support.**

**The numbers of A- and B-supporters will vary across election rounds.** The relative numbers of A- and B-supporters depends on which scenario applies. In each round, a random draw (carried out beforehand by a computer) determines which scenario applies. Between rounds, these draws are independent; the outcome for one round in no way affects the scenario drawn for another round. The likelihood of occurence indicates the chance that that scenario applies. In each round, there is a 40% chance that scenario 1 applies, a 50% chance that scenario 2 applies and a 10% chance that scenario 3 applies. This information is summarised in the table below.

| **Scenario** | **Number of active participants** | **Number of programmed A-supporters** | **Number of programmed B-supporters** | **Likelihood of occurence** |
|---|---|---|---|---|
| 1 | 3 | 6 | 0 | 40% |
| 2 | 3 | 4 | 2 | 50% |
| 3 | 3 | 2 | 4 | 10% |

In each election, your group of three active participants will be randomly reshuffled. It is therefore very unlikely that you will take part in the election with the same participants from one round to the next.

## B.3 Quiz 1

Please answer the following questions. You need to answer them correctly to access the next set of instructions and the main experiment.

- How many election rounds (excluding practice rounds) does the experiment have in total?
- How many voters in your electorate are programmed automatically?
- You will play the experiment with the same group of voters in every round. Is this correct?

## B.4 Part 2: Payoffs and Decisions

You will not be assigned the role of A- or B-supporter. Instead, your payoffs depend on the total number of votes received by the A- and B-candidates. The total number of votes is the sum of the votes cast by the programmed voters and the participants (including you). There will therefore be **9 votes in each election, 6 of which are cast automatically and 3 of which are cast by the participants.** To win the election, a candidate needs a majority of votes: this is **at least 5 of the 9 votes.**

In this experiment, you can earn experimental points. If the A-candidate obtains a majority of votes then the A-candidate wins the election, and you obtain **10 experimental points.** You can obtain additional points if the B-candidate obtains a certain number of the votes cast. If the B-candidate obtains 3 or 4 of the 9 votes then you will receive an additional 10 experimental points, earning you **20 experimental points**. However, if the B-candidate obtains a majority of votes, then you will receive **0 experimental points**. Your payoffs are summarised in the table below:

| Total votes for the A-candidate | Total votes for the B-candidate | Payoffs in experimental points |
|---|---|---|
| 0-4 | 5-9 | 0 |
| 5-6 | 3-4 | 20 |
| 7-9 | 0-2 | 10 |

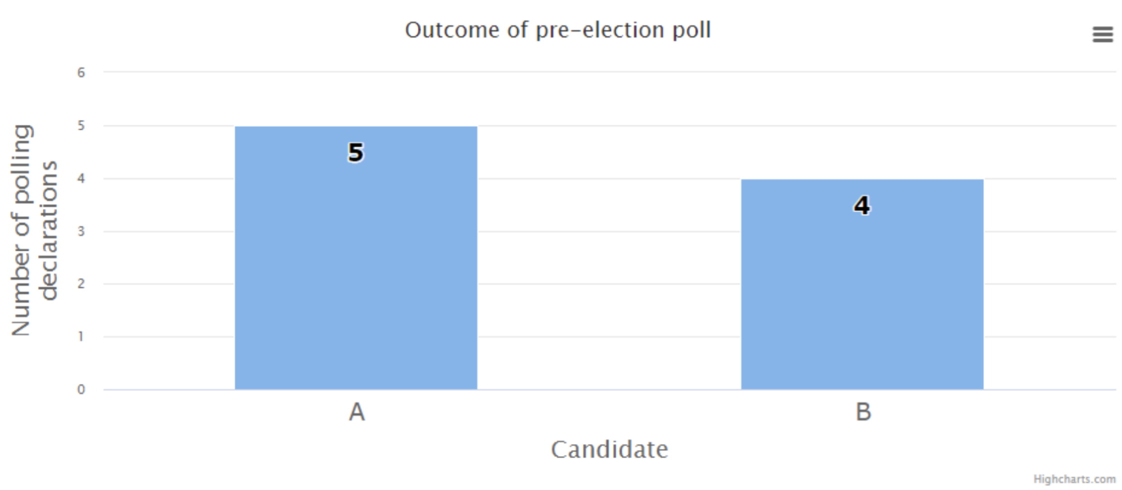
Before making your voting decision, you will take part in a pre-election poll. For this pre-election poll, you will be asked to declare which candidate that you intend to vote for in the upcoming election. **The A- and B-supporters are automatically programmed to declare an intention to vote for the candidate that they support**. Therefore, the A-supporters declare an intention to vote for the A-candidate and the B-supporters declare an intention to vote for the B-candidate. The declared intentions are then aggregated at the electorate level. You will be told how many polling declarations (including your own) were cast for each candidate and be shown a graph of the result. An example of that graph can be seen below:

After viewing the poll result, you will cast your vote in the election. You are free to deviate from your own polling declaration in the vote. After all votes have been counted, you will be informed of the result and how many experimental points you earned in that round. You will firstly play 5 practice rounds, then 20 main rounds of the experiment.

Your final payment is made up of **three components**:

1. It depends on the **number of experimental points you accumulate across 6 randomly chosen rounds** of the experiment. The conversion rate is such that **one point is equal to €0.10** so you receive a euro for every 10 points earned in the chosen rounds.

2. We will ask you a question about what you believe to be the actual composition of the electorate. **Your points from 3 rounds of this beliefs task** will contribute towards your



final payoff. These 3 rounds are selected at random and will be different from voting rounds that are used to determine your payoffs (as in item 1). This task will be explained in the third part of the instructions.

3. **You will receive a participation fee of €8** if you complete the experiment. Contrary to most other experiments, you may choose to try to increase this amount by investing in a gamble. If you do, **this amount may end up being higher or lower than €8**. This task will be explained after the main experiment.

Total payoffs are therefore given by:

$$\text{€ } \textit{total payoff} = (\textit{points from 6 rounds of the experiment} + \textit{points from 3 beliefs tasks}) \times 0.10 + \text{€ } \textit{amount earned in investment task}$$

## B.5 Quiz 2

Please answer the following questions. You need to answer them correctly to access the main experiment. To help you, the payoff table can be seen below.

- How many euro cent is each experimental point worth?
- Suppose that there are 4 A-supporters and 2 B-supporters. Also suppose that the 2 other participants vote for A. How many points do you earn if you vote for B?
- Again, suppose that there are 4 A-supporters and 2 B-supporters. But now suppose that the 2 other participants vote for B. How many points do you earn if you vote for B?

## B.6 Part 3: Beliefs task

In addition to the elections, you will take part in an additional task for which you can earn extra points. This takes place in every other round. In the even numbered rounds, you will see an additional page. You see this before you cast your vote. This page shows the poll result again and will ask you to provide an answer to the question:

***How likely do you think it is that there ARE two B-supporters in the electorate (i.e. that scenario 2 applies)?***

**You give your answer as a percentage, which indicates your belief that there ARE two B-supporters in the electorate**, that is, that scenario 2 applies. To determine your payment, the computer will randomly draw two numbers. For each draw, all numbers between 0 and 100 are equally likely to be selected (including decimals). Draws are independent; the outcome of the first draw in no way affects the outcome of the second.

**Your answer can go from** 0% (meaning that you are completely certain that there ARE NOT two B-supporters) **to** 100% (meaning that you are completely certain that there ARE two B-supporters). **Your earnings can be either 10 or 0 experimental points**. You win 10 experimental points if either of the following statements apply:

- If there ARE two B-supporters and the percentage you picked is larger than at least one of the two draws.
- If there ARE NOT two B-supporters and the percentage you picked is smaller than at least one of the two draws.

The chance of earning 10 experimental points therefore depends on two things:

1. The actual outcome (whether there are two B-supporters or not).
2. The percentage you selected as the answer to the question in bold above.

**You maximise your chance of earning 10 points if you report your beliefs as accurately as possible**. There is nothing to be gained by stating a percentage which differs to what you actually believe. For example, if it turns out that there are two B-supporters, the chance that you earn 10 experimental points increases the closer the likelihood you selected is to 100%. On the other hand, if it turns out that there ARE NOT two B-supporters, the chance that you earn 10 experimental points increases the closer your selected likelihood is to 0%. Moving away from your actual beliefs lowers your expected payoff.

You will be asked to enter your likelihood as a percentage. To help you understand the consequences of your answer, we will provide a calculator that allows you to see the chance (expressed as a percentage) that you win 10 experimental points. A screenshot of that calculator can be seen below for a hypothetical example of a likelihood of 75%. As the calculator shows for this percentage likelihood, if there ARE two B-supporters, you have a 94% chance of winning 10 experimental points. If there ARE NOT two B-supporters, you have a 44% chance of winning 10 experimental points.

After you enter your beliefs, you will not be told which scenario applies or what your payoff is. Your payoffs from the beliefs tasks will be included in your final payoff.

The following page will test your understanding of this task.

You can use the calculator below to test your chances of winning 10 experimental points. Please enter your likelihood percentage as a whole number (the calculator will round decimals down). The percentages are rounded to the nearest whole number.

| **Reported likelihood percentage (%)** | **Chance of winning if there ARE two B-supporters (i.e. scenario 2 applies)** | **Chance of winning if there ARE NOT two B-supporters (i.e. scenario 2 does not apply)** |
|---|---|---|
| 75 | 94 | 44 |

75 % Send

## B.7 Quiz 3

Please answer the following questions. You need to answer them correctly to access the main experiment. To help you, the payoff table can be seen below.

- If you provide a likelihood percentage of 100% and there ARE NOT two B-supporters (although you do not know this), what is the probability (expressed as a percentage) that you win 10 experimental points?
- If you provide a likelihood percentage of 20% and there ARE two B-supporters (although you do not know this), what is the probability (expressed as a percentage) that you win 10 experimental points?

## B.8 Summary of instructions

The experimenter will now provide you with a handout.

**Practice rounds**

Before starting the main experiment, you will play 5 practice rounds. Any experimental points earned from these rounds will not count towards your final payoff. After completing the practice rounds, you will proceed to the main experiment.

**Main rounds**

After the 5 practice rounds, you will take part in 20 main rounds. Both the main rounds and practice rounds include the beliefs task, which happens every other round.

**Investment task**

Once you have finished the main experiment, you will be asked a short question on an investment decision, in which you can choose to invest some of your participation fee. This will be explained later on during this session.

**Survey Questions**

We will then ask you a series of short survey questions.

**Payment**

After everybody has completed the experiment, you will be informed of your total payment for this session. You are then free to leave the experiment and your payment will be handed to you as you leave.

**Next steps**

Please press Next to proceed with the 5 practice rounds.

# C Figures

Figure C1: Rate of protest voting by round of the experiment.

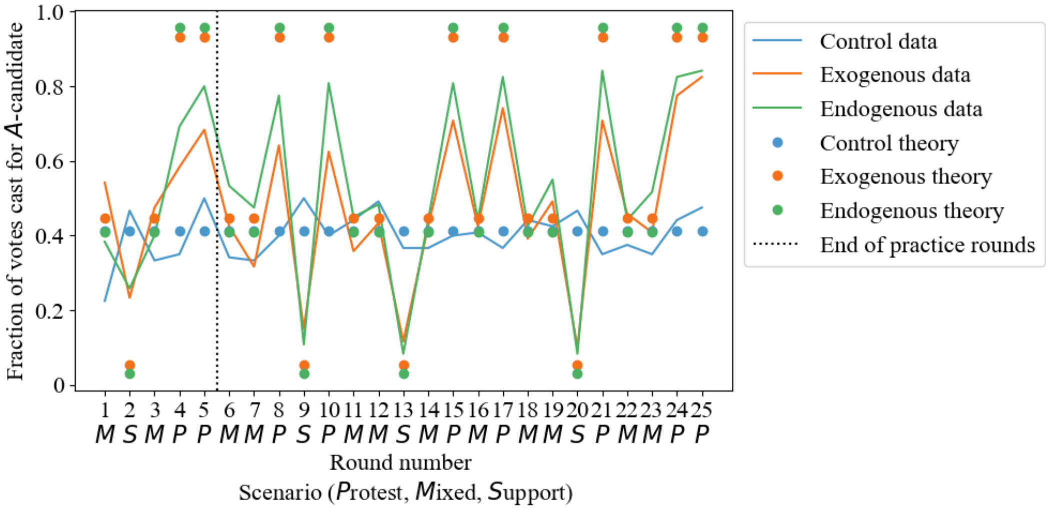
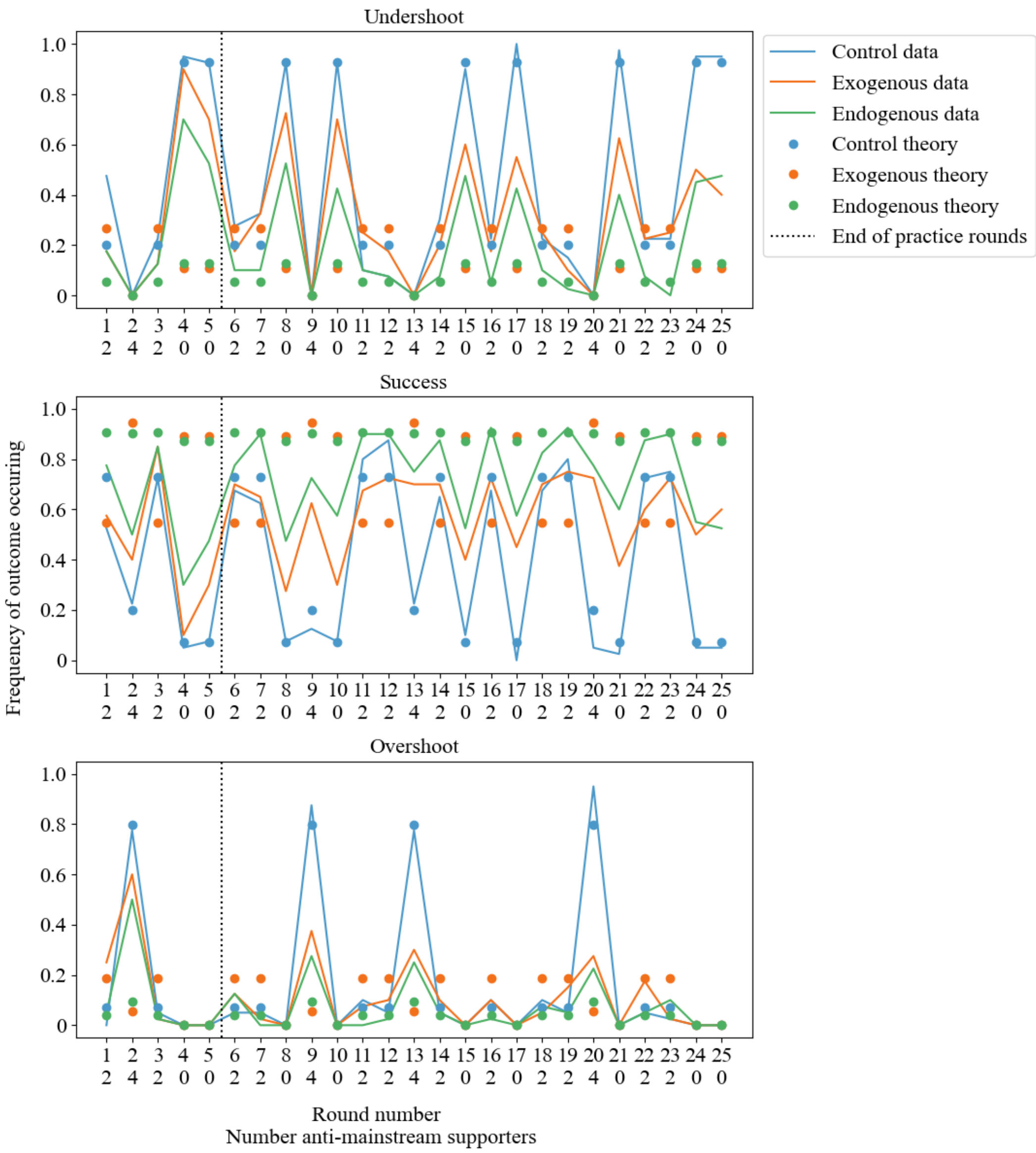


Figure C2: Rate of different outcomes by round of the experiment.




Note: We plot both theoretical predictions and empirical data for the likelihood that the protest undershoots, succeeds and overshoots across all 25 (so including practice rounds) of the experiment. Statistics from all three treatments are plotted.

# D Tables

Table D1: Theoretical predictions for protest voting, vote outcomes and beliefs by poll outcome and treatment

| **d-value** | **Treatment** | **Pr(protest vote)** | | | **Pr(outcome)** | | | **Beliefs** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $Pr(d\|\sigma)$ | | Overall | $d_i = 0$ | $d_i = 1$ | Undershoot | Success | Overshoot | Overall | $d_i = 0$ | $d_i = 1$ |
| 0, 1 | Exogenous | 1.000 | | | 0.000 | 1.000 | *not possible* | 0.000 | | |
| 0.206 | Endogenous | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 | *not possible* | 0.000 | 0.000 | 0.000 |
| 2 | Exogenous | 1.000 | | | 0.000 | 0.688 | 0.312 | 0.312 | | |
| 0.213 | Endogenous | 0.812 | 0.654 | 1.000 | 0.251 | 0.662 | 0.087 | 0.312 | 0.577 | 0.000 |
| 3 | Exogenous | 0.430 | | | 0.330 | 0.606 | 0.064 | 0.803 | | |
| 0.238 | Endogenous | 0.464 | 0.000 | 1.000 | 0.000 | 1.000 | 0.000 | 0.803 | 1.000 | 0.577 |
| 4 | Exogenous | 0.395 | | | 0.206 | 0.684 | 0.110 | 0.932 | | |
| 0.197 | Endogenous | 0.622 | 0.000 | 1.000 | 0.000 | 1.000 | 0.000 | 0.932 | 0.821 | 1.000 |
| 5 | Exogenous | 0.000 | | | 0.605 | 0.395 | 0.000 | 0.605 | | |
| 0.097 | Endogenous | 0.185 | 0.000 | 0.251 | 0.254 | 0.637 | 0.109 | 0.605 | 0.000 | 0.821 |
| 6, 7 | Exogenous | 0.000 | | | *not possible* | 1.000 | 0.000 | 0.000 | | |
| 0.048 | Endogenous | 0.000 | 0.000 | 0.000 | *not possible* | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Overall** | Control | 0.414 | | | 0.472 | 0.412 | 0.115 | 0.500 | | |
| **(weighted)** | Exogenous | 0.600 | | | 0.178 | 0.719 | 0.104 | 0.500 | | |
| | Endogenous | 0.631 | 0.346 | 0.826 | 0.078 | 0.893 | 0.029 | 0.500 | 0.523 | 0.414 |

Notes: The predicted likelihood for polling for A ($\sigma$) is 0.489. Theoretical predictions for the polls treatments are broken down by the total number of polls cast for the anti-mainstream (the value of $d$). Below the value of $d$ we present the predicted likelihood with which each value of $d$ occurs.

Table D2: Results for protest voting and vote outcomes by poll outcome and treatment

| **d-value** | **Treatment** | **Pr(protest vote)** | | | **Pr(outcome)** | | | **Beliefs** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $Pr(d\|\sigma)$ | | Overall | $d_i = 0$ | $d_i = 1$ | Undershoot | Success | Overshoot | Overall | $d_i = 0$ | $d_i = 1$ |
| 0, 1 | Exogenous | 0.753 | | | 0.556 | 0.444 | *not possible* | 0.437 | | |
| 0.239 | Endogenous | 0.811 | 0.784 | 0.910 | 0.461 | 0.539 | *not possible* | 0.310 | 0.320 | 0.270 |
| 2 | Exogenous | 0.616 | | | 0.353 | 0.525 | 0.122 | 0.661 | | |
| 0.206 | Endogenous | 0.674 | 0.554 | 0.911 | 0.318 | 0.624 | 0.058 | 0.586 | 0.643 | 0.497 |
| 3 | Exogenous | 0.476 | | | 0.161 | 0.765 | 0.074 | 0.637 | | |
| 0.245 | Endogenous | 0.458 | 0.169 | 0.940 | 0.045 | 0.922 | 0.051 | 0.617 | 0.549 | 0.738 |
| 4 | Exogenous | 0.221 | | | 0.343 | 0.534 | 0.123 | 0.647 | | |
| 0.171 | Endogenous | 0.427 | 0.035 | 0.795 | 0.027 | 0.922 | 0.051 | 0.657 | 0.549 | 0.738 |
| 5 | Exogenous | 0.092 | | | 0.171 | 0.651 | 0.178 | 0.654 | | |
| 0.083 | Endogenous | 0.124 | 0.081 | 0.182 | 0.117 | 0.647 | 0.236 | 0.565 | 0.476 | 0.693 |
| 6, 7 | Exogenous | 0.108 | | | *not possible* | 0.675 | 0.325 | 0.576 | | |
| 0.056 | Endogenous | 0.123 | 0.122 | 0.128 | *not possible* | 0.651 | 0.349 | 0.494 | 0.684 | 0.421 |
| **Overall** | Control | 0.407 | | | 0.438 | 0.401 | 0.161 | 0.556 | | |
| **(weighted)** | Exogenous | 0.476 | | | 0.311 | 0.595 | 0.094 | 0.608 | | |
| | Endogenous | 0.538 | 0.417 | 0.736 | 0.191 | 0.745 | 0.064 | 0.551 | 0.530 | 0.604 |

Notes: The observerd likelihood of polling for A ($\sigma$) is 0.380. Observed protest votes and outcomes are broken down by the total number of polls cast for the anti-mainstream (the value of $d$). Below the value of $d$ we present the frequency with which each value of $d$ occurs.