Information, Closeness of Elections and Authoritarianism

Dominic Rohner
University of Cambridge

December 31, 2006

Abstract

Often democracy is defeated by democratically elected leaders who pursue undemocratic policies. The present contribution seeks to explain how information can affect the danger of incompetent or populist politicians being elected into office. It is shown with the help of a voting model that both the proportion of informed voters in a society and the state of the economy matter for determining who gets elected and with what margin of victory. It is argued that a high proportion of uninformed voters can result in massive popular support for a "bad" party, which can lead to authoritarianism.

JEL Classification: D72, D83, H11.

Keywords: Voting, Closeness of Elections, Information, Democracy, Authoritarianism.

1 Introduction

Democracy is often associated with many desirable outcomes such as economic prosperity, civil liberties, and good governance[1]. Unfortunately, as highly prised as it is, democracy has often proved to be fragile and reversials of democratisation have thrown many countries in political chaos and economic misery. Paradoxically, democracy is not always abolished by a coup, but often threatened by democratically elected "bad" politicians who pursue incompetent, authoritarian and populist policies. Examples do not only include the infamous election of Adolf Hitler in Germany in 1933, but also the electoral successes in recent decades of undemocratic and populist politicians and parties such as Robert Mugabe in Zimbabwe, Hugo Chavez in Venezuela and Hamas in Palestine. Thus, it seems of foremost importance to know under what conditions democracy can persist and does not defeat itself by electing "bad" leaders.

[1] *Address:* Faculty of economics, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom. dr296@cam.ac.uk.

Acknowledgements: I would like to thank Partha Dasgupta, David Myatt, Scott Page, Hamid Sabourian, Peyton Young and seminar participants in Cambridge for helpful comments.

The present contribution assesses with the help of a voting model the impact of information on the likelihood of a "bad" politician being elected into office. It will be shown that the proportion of uninformed voters in a society affects both the probability of a "bad" politician being elected, as well as the closeness of the election.

The small formal literature about consolidation of democracy and bad leaders is relevant for the present paper. Acemoglu and Robinson (2001, 2006) study the consolidation of democracy by focussing on the risk of violent transitions ("coups"), which are linked to the level of inequality. They provide important insights about the elite weakening democracy by plotting coups if democracy is too redistributive, but do not address the case of democracy being defeated by democratically elected leaders. Caselli and Morelli (2004) analyse how "bad politicians" can emerge in the framework of a citizen-candidate game. Contrary to our paper they focus on the entry into politics rather than on voters making wrong choices.

The present paper is also related to the literature on voting. Recently, many interesting papers focussed on pivotal and strategic voting (Austen-Smith and Banks, 1996; Feddersen and Pesendorfer, 1996; 1997; Myatt, 2006). The articles which are most relevant for the present contribution link voting with information issues. Two recent papers by Martinelli (2006) and Gul and Pesendorfer (2006) are devoted to this topic. They provide important insights, but focus on other issues than the present paper. Martinelli (2006) assesses in a theoretical framework under what conditions voters would be "rationally ignorant" or would acquire information. The size of the electorate and the marginal cost of information are decisive parameters. Gul and Pesendorfer (2006) show how in a framework with ignorant voters candidates with a "preferred personality" can choose partisan politics and get re-elected although their policies imply a lower payoff for the electorate. They focus on explaining why the Downsian policy convergence result does not necessarily hold with charismatic candidates. By contrast, in the present contribution no candidate has a "personality advantage", and the state of the economy influences voter decisions. Also, the emphasis of the present paper lies on the effects of the composition of the electorate (informed versus non-informed voters) and the margin of victory rather than the implication of "personality preferences".

Another literature that is of interest for the present contribution is the one studying the causes and effects of dictatorships (see for example Moore, 1993; Olson 1993; Wintrobe, 1998).

The remainder of the paper is organised as follows: In section 2 the voting game is described, and section 3 emphasises the link between the closeness of elections and authoritarian policies. Section 4 concludes. In Appendix A the model's results are discussed for pivotal voting. In Appendix B a simple model is built that allows for an endogenous determination of the proportion of informed voters.

2 The Model

The present section develops a simple voting game where a part of the voters are informed about the relative quality of the two competing politicians/parties[2] (one left-wing versus one right-wing) and the rest are uninformed. The attributes "bad" and "good" are used for describing the competing politicians. "Bad" can have two meanings: first, "bad" can mean incompetent; second, "bad" can mean immoral, in the sense of undemocratic and authoritarian. The model we build allows for both meanings, and history has taught us that they often go together. An economy can only perform well if a certain level of freedom is guaranteed, and it is very rare that economic freedom can exist in the long-run without political freedom. In the vast majority of cases, immoral dictatorial rulers make choices that hinder economic development, and can therefore be regarded as incompetent from an economic perspective. In the remainder of the paper the attributes "bad", "incompetent" and "authoritarian" are used interchangeably.

The uninformed voters face a somewhat informative, but noisy signal that they take into account for their voting decision. When it comes to choosing the appropriative party, the voters care about both the level of competence and the level of redistribution favoured by the two parties. It will be shown what impact information has on the closeness of election and on the risk of an incompetent party being elected into office. The following subsection is devoted to a discussion of the assumptions.

2.1 Assumptions

G.1 Utility function. The utility function is of the following form for a given voter j:

$$U_j = \theta v_w \pi_w + (1-\theta) v_l \pi_l \tag{1}$$

where v=share votes, π=payoff from a given party, θ=parameter ($0.5 \leq \theta \leq 1$), subscript w=winning party, subscript l=loosing party.

The utility function introduced in equation (1) allows for different weights given to the looser's policy stand and vote share in the utility function. If $\theta = 0.5$, voters do care about the votes of the elected party and the opposition to an equal extend. This corresponds to the case of "sincere voting" often emphasised in the literature. For countries with strong blockage possibilities for the opposition, this parameter value is reasonable. By contrast, if $\theta = 1$, voters do only care about the winner. This corresponds to the case emphasized by the pivotal voting literature, where voters do Bayesian updating and focus exclusively on the optimal decision for the event in which they are pivotal. The margin of

[2] In the present framework there are two parties with one leader each. The implications of the model hold for both cases of competing politicians and of competition between parties. Therefore, the expressions "parties", "politicians", "candidates" and "leaders" are used interchangeably in the paper.

victory is not taken into account. This case has received considerable attention recently (e.g. Austen-Smith and Banks, 1996; Feddersen and Pesendorfer, 1996; 1997), but has also been strongly critisised for using non realistic assumptions (Margolis, 2001; 2002). Indeed, it has been established in several papers that theoretically and empirically the probability of being pivotal in an election is extremely small (Gelman, King, and Boscardin, 1998; Fischer, 1999; Mulligan and Hunter, 2003). Gelman, King, and Boscardin (1998), for example, find that the probability of a vote being pivotal in a close US presidential election, such as the one of 1992, is about 1 in 10 millions. It is difficult to believe that such small probabilities play the only role in the voter's considerations, as assumed by the pivotal voting literature.

Therefore, it is reasonable to think that voters do not only care about the tiny probability of being pivotal, but that they also take the magnitude of the margin of victory into consideration. In this way many phenomena, such as people voting for parties that are given no chance by opinion polls of winning the election, can be explained. Such phenomena are not compatible with a framework of pure pivotal voting.

Voters not exclusively focussing on being pivotal translates into the assumption of $\theta < 1$. In this case, the political agenda and the vote share of the non-elected party enter the utility function as well. This reflects the blockage power and present and future political influence of the opposition party, which increase in its vote share. This blockage power is the main reason why the closeness of elections matters[3] and plays an important role as far as the risk of

[3] Historical real-world evidence suggests that it makes sense to include the political platform and vote share of the opposition in the utility function. Several countries have seen the big left- and right-wing parties sharing power and forming "grand coalitions". In many consensual countries (using the concept and terminology of Lijphart, 1999) such power-sharing has even been institutionalised. In Switzerland, for example, the big right-wing and left-wing parties form together the government and their relative vote share determines the number of ministers they obtain. Until recently, the number of minister posts where assigned quasi-fixely to the parties according to the so-called "magic formula". However, vote shares still mattered heavily for intra-governmental bargaining. At present the vote shares directly determine the number of ministers obtained. Similar cases include Belgium, the Netherlands and Scandinavian countries.

Also in the big European countries such as Germany or France power-sharing has been observed. Germany has had several Grand Coalitions between the two big parties CDU/CSU and SPD in its history. The latest has started in 2005 under the leadership of the Bundeskanzlerin Angela Merkel. In France, when electoral support was very divided, "cohabitation" took place three times so far with the President belonging to one block and the Prime Minister belonging to the other block. The latest case was between 1997 and 2002 with Jacques Chirac as President and Lionel Jospin as Prime Minister.

What has also frequently been observed in history is the dominance of one party or party coalition, and elections determining the power balance inside the governing coalition. This has been the case in recent decades for Japan's Liberal party, for Singapore's ruling People's Action Party or for Italy's Democrazia Christiana until the beginning of the 1990s. Our utility function can also account for this kind of situation.

However, also in countries with a pure two party system such as the United States, or the United Kingdom (if one ignores smaller parties such as the Liberal Democrats) the vote shares matter, and the winner "does not take it all" *stricto senso*. Tony Blair, for example, faces the strong resistance of the shadow cabinet of the Tories in whatever enterprise he initiates. One could clearly observe after the three election victories of Labour under Blair how much the

authoritarianism is concerned.

For the rest of this section it will be assumed that voters care to an equal extent about vote shares of the opposition and of the government. This assumption simplifies the analysis, and facilitates an intuitive understanding of results that hold as well for more general values of $0.5 \leq \theta < 1$. The results for those intermediate cases $0.5 \leq \theta < 1$ and for purely pivotal decision making are treated in Appendix A.

The payoff functions for the two parties are an important part of the utility function. Without loss of generality, the payoff of the incumbent can be normalised to $\pi_G = 0$. The payoff from voting for the challenger is displayed in equation (2).

$$\pi_C = S + E \tag{2}$$

where S=variable related to the net gains/losses from redistribution for voting for the challenger, E=variable related to the relative competence ("efficiency") of the challenger, subscript G=incumbent, subscript C=challenger.

The variable S is uniformly distributed between -1 and 1. It takes its minimum value S=-1 for the voter who looses most from the redistribution policy of the challenger as compared to the redistribution policy of the incumbent. S takes the value of S=1 for the voter who gains most. If the incumbent would be, say, the left-wing party, then the richest voter in the economy would gain most from an election of the right-wing challenger and would accordingly have a value of S=1. On contrary, the poorest voter would have S=-1, the "median" voter S=0. For the incumbent being right-wing it would be just the other way around. The uniform distribution reflects the assumption that the level of wealth is continuously distributed in the society[4].

Both the incumbent and the challenger can either be competent or incompetent. Incompetent parties have $e_j = 0$, competent parties have $e_j = e$, where j=C,G, $0 < e < 1$. As the payoff from voting for the incumbent is normalised to zero, the voting decision depends on the competence difference between the candidates. This is captured by the variable $E = e_C - e_G$, which can take three values, i.e. $E \in \{-e, 0, e\}$, where $e = -(-e),\ 0 < e < 1$. The parameter value of E=(-e) indicates that the challenger is relatively less competent than the incumbent (thus, $e_C = 0, e_G = e$), E=0 means that they have the same level of competence (either $e_C = 0, e_G = 0$ or $e_C = e, e_G = e$), E=e corresponds to the challenger being more competent ($e_C = e, e_G = 0$).

closeness of election affects the margin of manoeuvre of the government. After the third, more close election, Tony Blair has less *de facto* political power due to stronger political support for the opposition.

[4]Relaxing this simplifying assumption would not alter the main results of the model. As long as the distribution has a symmetrical shape and a median of 0 the results would remain the same. A skewed distribution or a median above or below 0 would introduce a bias in favour of one of the parties, but would not affect the fundamental results and implications of the model.

The informed voters observe both S and E, the uniformed observe only S, but not E. They do, however, receive a noisy signal about E, which will be discussed under G.3.

G.2 Actions. The voters maximise the value of their utility function (1) by choosing between voting for the incumbent (G) or for the challenger (C). The party which gets more votes is elected in office, a tie is broken with equal probabilities.

G.3 Information. Both parties have *a priori* equal probabilities to be competent or not, i.e. with probability 0.5 they have $e_j = 0$, with probability 0.5 they have $e_j = e$, where j=C,G. If informed voters can directly observe the level of e_j for both parties, and thus compute the relative efficiency $E = e_C - e_G$ of the challenger, the uniformed voters observe a noisy signal Y, which refers to the state of the economy. Having a competent government is neither a necessary nor a sufficient condition for a good economic outcome, but economic success is relatively more likely under a competent rather than an incompetent government. This makes the economic state of the country a noisy, but somewhat informative signal about the level of competence of the government. For simplicity, two stylised states of the economy are distinguished: Y^+=boom and Y^-=recession[5]. The probability that under a "good" incumbent the economy is doing well is given by $prob(Y^+ \mid e_G = e) = a$, where $0.5 < a < 1$. It follows that $prob(Y^- \mid e_G = e) = (1 - a)$. It is assumed that the odds for economic success are worse for a "bad" incumbent: $prob(Y^+ \mid e_G = 0) = b$ and $prob(Y^- \mid e_G = 0) = (1 - b)$, where $0 < b < 0.5$. These probabilities are common knowledge and are used by the voters for updating their beliefs about the level of competence of the incumbent. By contrast, the economic situation does not tell the voters anything about the ability level of the challenger.

2.2 Intermediate Results and the Equilibrium Solution

For computing the equilibrium outcome of the game, we need to work out intermediate results for all possible cases.

2.2.1 Informed voters

Informed voters can observe both the level of competence of the competing parties and the difference in redistribution policies. A voter elects the challenger if the condition below holds (without loss of generality we can normalise $\pi_G = 0$):

$$\pi_C^i > \pi_G^i \Leftrightarrow S + E > 0 \tag{3}$$

where subscript i refers to informed voters.

For E=(-e), which corresponds to the case when the incumbent is more competent than the challenger, only voters who could win a lot from a switch in redistribution policy vote for the challenger. If for example the challenger was

[5] The results are robust to the inclusion of more that two states of the economy.

the right-wing party, very rich voters (with S>e) would still vote for that party despite its incompetence, as they could save much money with lower taxes. The vote shares of the two parties for E=(-e) are displayed below.

$$(v_C^i \mid E = -e) = \frac{1-e}{2},\ (v_G^i \mid E = -e) = \frac{1+e}{2} \tag{4}$$

where v_C^i=share votes from informed voters for the challenger, v_G^i=share votes from informed voters for the incumbent.

By contrast, for E=0, informed voters decide on purely economic grounds and choose the party which is closer to their preferred redistribution policy. Thus, the vote shares become as follows.

$$(v_C^i \mid E = 0) = \frac{1}{2},\ (v_G^i \mid E = 0) = \frac{1}{2} \tag{5}$$

For E=e, most voters prefer the challenger, with the exception of those who have too much to loose from altered redistribution policies.

$$(v_C^i \mid E = e) = \frac{1+e}{2},\ (v_G^i \mid E = e) = \frac{1-e}{2} \tag{6}$$

If a population is composed by only informed voters, the least competent party never gets elected and electoral outcomes correspond to the social optimum.

The margin of victory is defined as the absolute value of the difference between the vote share of the challenger and half of the votes.

$$M = |v_C - 0.5| \tag{7}$$

where M=margin of victory.

Whenever one party is more competent than the other, under full information the margin of victory becomes $M = \frac{e}{2}$. If there is a difference in competence between the two parties, the most competent one will be elected. If the level of competence is the same, the margin of victory becomes M=0. This is summarised below.

Lemma 1 *For a population entirely composed by informed voters either the more competent party is elected with a margin of victory of $M = \frac{e}{2}$, or if both are (in-)competent, there is a very close election with M=0, and each party gets elected with equal probability.*

Proof. Follows from the discussion above. ■

This result is important for the further analysis, as the least desirable outcome for a society in terms of authoritarianism is an incompetent party being elected with massive popular support. This is avoided if all voters are informed.

2.2.2 Uninformed voters

The analysis is different for uninformed voters, who do not directly observe the competence of parties, but only the current state of the economy.

The uninformed voters do Bayesian updating and form their beliefs about the competence level of the incumbent based on the state of the economy. For an economic boom, Y^+, the beliefs about the probability that the incumbent is "good" are given by $\mu = prob(e_G = e \mid Y^+) = \frac{0.5a}{0.5a+0.5b} = \frac{a}{a+b} = p$. As $a > b$, it must be that $p > \frac{1}{2}$. The incumbent is believed to be "bad" with probability $(1-\mu) = (1-p)$. Similarly, for observing Y^- the beliefs are $\mu = prob(e_G = e \mid Y^-) = \frac{0.5(1-a)}{0.5(1-a)+0.5(1-b)} = \frac{(1-a)}{(1-a)+(1-b)} = (1-w)$ and $(1-\mu) = w$ (this notation facilitates future comparisons). As $a > b$, it must be that $w > \frac{1}{2}$.

Given that the state of the economy does not reveal any information about the ability level of the challenger, the beliefs are $\mu = prob(e_C = e \mid Y^+) = 0.5$ and $\mu = prob(e_C = e \mid Y^-) = 0.5$.

Given these probabilities, one can compute the expected values of E displayed below:

$$prob(E = -e \mid Y^+) = \frac{p}{2},\ prob(E = 0 \mid Y^+) = \frac{1}{2},\ prob(E = e \mid Y^+) = \frac{1-p}{2} \tag{8}$$

$$prob(E = -e \mid Y^-) = \frac{1-w}{2},\ prob(E = 0 \mid Y^-) = \frac{1}{2},\ prob(E = e \mid Y^-) = \frac{w}{2} \tag{9}$$

For the state of the economy being Y^+, the expected value of voting for the challenger is displayed below:

$$(\pi^u_C \mid Y^+) = S + \frac{p}{2}(-e) + \frac{1}{2}0 + \frac{1-p}{2}e \tag{10}$$

where superscript u refers to uninformed voters.

Knowing that (by normalisation) $\pi^u_G = 0$, we have

$$(\pi^u_C \mid Y^+) > \pi^u_G \Leftrightarrow S > (p - \frac{1}{2})e \tag{11}$$

As displayed in equation (12), the vote share of the challenger, v^u_C, is below 0.5[6]. This is intuitive, as the economy being above the long-run equilibrium path indicates that the incumbent is likely to be competent.

The model predicts that for a state of the economy of Y^+ most of the uniformed voters support the incumbent. This implication of the model is in line

[6] The vote share of the incumbent, v^u_G, equals $(1 - v^u_C)$. For convenience, it will only be referred to v^u_C in the remainder of the paper.

with the empirical finding that the incumbent has better re-election prospects if the economy is in a good shape (Fair, 1978; Brender and Drazen, 2005). Interestingly, Brender and Drazen (2005) find that the positive impact of growth on the probability of re-election is strongest in less developed countries and new democracies. It is reasonable to think that these kinds of countries have a higher share of uninformed voters. Thus, Brender and Drazen's findings are consistent with our model's prediction in subsection 2.2.3 that especially for countries with a high proportion of uninformed voters the economic situation is decisive in elections.

$$(v_C^u \mid Y^+) = \frac{1-(p-\frac{1}{2})e}{2} \tag{12}$$

The reasoning for Y^- is exactly symmetrical to the one for Y^+. Thus, we obtain:

$$(\pi_C^u \mid Y^-) > \pi_G^u \Leftrightarrow S > -(w-\frac{1}{2})e \tag{13}$$

$$(v_C^u \mid Y^-) = \frac{1+(w-\frac{1}{2})e}{2} \tag{14}$$

For an uninformed population the closeness of election and welfare implications of voting are different than for informed voters. Lemma 2 summarises the results.

Lemma 2 *For a population entirely composed by uninformed voters and for signals Y^+ (Y^-), the incumbent (challenger) gets elected and the margin of victory equals $M = \frac{(p-\frac{1}{2})e}{2}$ ($M = \frac{(w-\frac{1}{2})e}{2}$).*

Proof. For Y^+, $M = |(v_C^u \mid Y^+) - 0.5| = \frac{(p-\frac{1}{2})e}{2}$, for Y^-, $M = |(v_C^u \mid Y^-) - 0.5| = \frac{(w-\frac{1}{2})e}{2}$. ■

However, the state of the economy is not perfectly correlated with the competence of the incumbent. With positive probability a competent incumbent is replaced with an incompetent challenger if the incumbent's ability is not reflected in good economic outcomes. Similarly, with positive probability an incompetent incumbent obtains more votes than a competent challenger. Lemma 3 states under what conditions such socially suboptimal choices take place.

Lemma 3 *For E=(-e) (i.e., $e_C = 0, e_G = e$), with probability $(1-a)$ the (less competent) challenger will be elected, and for E=e (i.e., $e_C = e, e_G = 0$), with probability b the (less competent) incumbent will be re-elected, which both result in a socially non-optimal outcome.*

Proof. The probability of a less competent challenger being elected equals $prob(Y^- \mid E = -e) = prob(Y^- \mid e_G = e, e_C = 0) = (1-a)$. The probability of a less competent incumbent being re-elected is given by $prob(Y^+ \mid E = e) = prob(Y^+ \mid e_G = 0, e_C = e) = b$. ■

Although on average uninformed voters make the "correct" decision, they vote for the "wrong" candidate with positive probability. This result of lack of information leading to voting mistakes is in line with empirical findings (Heath, Andersen and Sinnott, 2002).

2.2.3 Population with a proportion q of voters being informed

At present we discuss the case of a population with proportion q of voters being informed (where 0<q<1), and accordingly proportion (1-q) being uninformed. For a state of the world E=(-e), the expected value of the share of votes obtained by the challenger can be computed, using equations (4), (12) and (14).

$$\begin{aligned}(v_C \quad | \quad E=-e) &= q\frac{1-e}{2} + (1-q)\left[a\frac{1-(p-\frac{1}{2})e}{2} + (1-a)\frac{1+(w-\frac{1}{2})e}{2}\right] \\ &= \frac{1}{2} - \frac{e}{2}\left[q + (1-q)\left[a(p-\frac{1}{2}) - (1-a)(w-\frac{1}{2})\right]\right] \qquad (15)\end{aligned}$$

As $a > (1-a)$, the expression $(v_C \mid E=-e)$ becomes smaller than $\frac{1}{2}$, provided that p and w are of a similar magnitude. For the rest of the analysis it will be assumed that this condition is fulfilled.

For E=(-e), the incumbent is more competent than the challenger, and gets on average more votes. The margin of majority equals $M = \left|(v_C \mid E=-e) - \frac{1}{2}\right| = \frac{e}{2}\left[q + (1-q)\left[a(p-\frac{1}{2}) - (1-a)(w-\frac{1}{2})\right]\right]$, which is increasing in q. Thus, for the incumbent being more competent, a higher fraction of the population being informed will result in more popular support for the incumbent and a less close election.

For E=0, two cases have to be distinguished. If $(e_C = e, e_G = e)$ holds, the vote share of the challenger becomes as displayed below:

$$(v_C \mid e_C = e, e_G = e) = \frac{1}{2} - \frac{e}{2}(1-q)\left[a(p-\frac{1}{2}) - (1-a)(w-\frac{1}{2})\right] \qquad (16)$$

For the usual assumption that p and w are of a similar magnitude, the vote share of the challenger is below 0.5. For a population with only informed voters the share of votes obtained by the challenger would have equaled exactly 0.5. The bias towards the incumbent in a mixed population is due to the high probability of an economic boom which induces uninformed voters to vote for the incumbent. The margin of victory, $M = \frac{e}{2}(1-q)\left[a(p-\frac{1}{2}) - (1-a)(w-\frac{1}{2})\right]$, is decreasing in the proportion of informed voters, q. In this case of the re-elected incumbent being competent, the bias in the margin of victory is less worrying that for the second case discussed below.

The second case, where $(e_C = 0, e_G = 0)$, is displayed below in equation (17).

$$(v_C \mid e_C = 0, e_G = 0) = \frac{1}{2} + \frac{e}{2}(1-q)\left[(1-b)(w-\frac{1}{2}) - b(p-\frac{1}{2})\right] \qquad (17)$$

Given that $(1-b) > b$, the challenger's vote share will be greater than 0.5 (assuming again that p and w are of a similar magnitude). The margin of victory, $M = \frac{e}{2}(1-q)\left[(1-b)(w-\frac{1}{2}) - b(p-\frac{1}{2})\right]$, is increasing in the number of uninformed voters. In this case the impact of the lack of information is more worrying, as a greater proportion of uninformed voters lead to a bigger margin of victory for the "bad" challenger.

If E=e, the reasoning is similar as before, and we receive the vote share for the challenger displayed in equation (18).

$$(v_C \mid E = e) = \frac{1}{2} + \frac{e}{2}\left[q + (1-q)\left[(1-b)(w-\frac{1}{2}) - b(p-\frac{1}{2})\right]\right] \tag{18}$$

We have $M = \frac{e}{2}\left[q + (1-q)\left[(1-b)(w-\frac{1}{2}) - b(p-\frac{1}{2})\right]\right]$, which is increasing in q. The insights of the discussion above lead to proposition 1.

Proposition 1 *For the case of a population with a share q of informed voters, the more competent candidate obtains on average a higher expected share of votes than her opponent. The level of expected popular support for the more competent candidate, and accordingly as well the margin of victory M increases in q.*

When both parties are of the same quality, the incumbent (challenger) wins the election for $e_G = e$ ($e_G = 0$). The margin of victory decreases in q.

Proof. Follows from the discussion above. ■

It is interesting what contrasting effects the proportion of informed voters, q, has on the margin of victory. If one candidate is more competent, q *increases* the margin of victory, while it *decreases* the margin of victory if both parties are of the same quality.

Although *on average* a mixed population with a share q of uninformed voters supports the more competent party, *sometimes* uninformed players vote for the less competent party. As mentioned earlier, uninformed voters cannot observe the competence of parties directly, but only a noisy signal, the state of the economy. Thus, as shown in lemma 3, uninformed voters support with positive probability the least competent of parties. This is the case if the incumbent is less competent, but the economy booms ($E = e, Y^+$), or if the challenger is less competent, but the economy is in a recession ($E = -e, Y^-$). The vote share for a competent challenger in a booming economy with an incompetent incumbent is displayed below.

$$\begin{aligned}(v_C \quad \mid \quad E = e,\ Y^+) &= q\frac{1+e}{2} + (1-q)\frac{1-(p-\frac{1}{2})e}{2} \\ &= \frac{1}{2} + \frac{e}{2}\left[q - (1-q)(p-\frac{1}{2})\right]\end{aligned} \tag{19}$$

We have $\frac{\partial(v_C|E=e,\,Y^+)}{\partial q} > 0$, indicating that the vote share for the challenger (who is more competent) is increasing in the percentage of informed voters in

the society. Whether the expression (19) is greater or smaller than 0.5 depends on the level of q. The critical level of q, labelled q^*, is displayed in equation (20) below.

$$(v_C \mid E = e,\ Y^+) > \frac{1}{2} \Leftrightarrow q > q^* = \frac{p - \frac{1}{2}}{p + \frac{1}{2}} \tag{20}$$

This is to say, if $q < q^*$ the incumbent gets re-elected even though the challenger is more competent. The reasoning is similar for the challenger being less competent than the incumbent and the economy being in a recession. In this case, the following two conditions hold.

$$(v_C \mid E = -e,\ Y^-) = \frac{1}{2} - \frac{e}{2}\left[q - (1-q)(w - \frac{1}{2})\right] \tag{21}$$

$$(v_C \mid E = -e,\ Y^-) < \frac{1}{2} \Leftrightarrow q > q^* = \frac{w - \frac{1}{2}}{w + \frac{1}{2}} \tag{22}$$

Again, if $q < q^*$, the society will choose the socially suboptimal party. As pointed out earlier, it does not only matter *which* party is elected, but also *how* it is elected. We can easily see that for $q < q^*$ (which implies that the incompetent challenger is elected), further decreases in q will increase the margin of victory of the "bad" party. Put differently, the lower the number of informed voters in a society, the higher is the popular support for the incompetent party. The results obtained above are summarised in proposition 2.

Proposition 2 *In a society with a positive share (1-q) of uninformed voters, an incompetent incumbent can be re-elected in place of a competent challenger if the economy is doing well, $(E = e, Y^+)$. Similarly, if the economy is in a recession, $(E = -e, Y^-)$, an incompetent challenger can win an election against a competent incumbent. These "wrong" electoral choices take place if there are only few informed voters, $q < q^*$. If the "bad" party is elected, its margin of victory increases for a higher proportion of uninformed voters.*

Proof. Follows from the discussion above. ■

Figure 1 illustrates the level of electoral support for a "bad" challenger in economically hard times, $(E = -e, Y^-)$, for the particular parameter values e=0.8 and w=0.9.

The findings of proposition 2 have important implications. Once the proportion of informed voters q is small enough, and the least competent party is elected, a further decrease in q increases the margin of majority and thus the political power of this undesirable party. As argued earlier, massive popular support for a "bad" party can result in abuse of power and authoritarian policies. The following section is devoted to this issue.

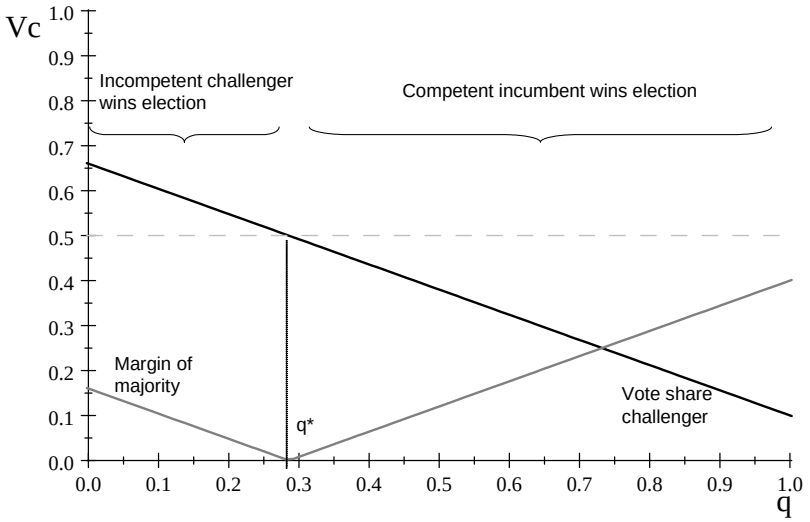

Figure 1: Electoral support for a "bad" challenger in a recession

3 Why the Margin of Victory Matters

In the framework of our model, massive support of a large part of the population for the left-wing or for the right-wing party leads to a large majority for the winner. A real-world example of a mostly uninformed and apolitical electoral body and of a "bad" challenger being elected in economically hard times with a massive popular support is the election of Adolf Hitler 1933 in Germany. Hitler received a huge power boost after about 90% of Austria's population voted for the "Anschluss" (the reunification with Nazi Germany) in 1938. Without this incredible popular support and the absence of a strong internal opposition, Hitler would not have been able to enforce subsequently his totalitarian policies. It is difficult to overstate the catastrophic consequences of the large popular support for Hitler, and of the totalitarian policies which it enabled.

Usually voters and political analysts are above all concerned with *who* is elected. What, however, matters as well heavily is *how*, i.e. with what margin of victory. It is essential for a democratic system that no "bad" party is omnipotent. Both in two-party Westminster-style democracies such as the United Kingdom as well as in consensual coalition-prone democracies such as Switzerland it is important that the power of the dominant party (or parties) is limited either by the opposition or by the coalition partners (see Lijphart, 1999, for a discussion of these two forms of democracy). Wherever the power of a "bad" party is non-balanced, the risk of abuse of power is almost inevitable, no matter what the actual content of the party doctrine is. This issue has been emphasised in Michels' (1962 [1911]) classic "iron law of oligarchy" and is well-illustrated

by the egalitarian utopia of soviet communism which turned in Stalinist terror.

More formally, we can make the assumptions that a) several key policies require a super-majority of r, whereas $0.5 \leq r \leq 1$, and b) that the opposition party blocks as many policy projects as it can. Under those assumptions a close election can prevent a "bad" party from implementing authoritarian policies.

Both assumptions a) and b) appear reasonable in the light of empirical evidence. In many countries amendments of the constitution involve acceptance by all chambers of the Parliament, often with a qualified majority. In some countries, such as for example Switzerland, even a popular vote is required for important decisions. But also in countries where only a simple parliamentary majority is needed, the government usually needs a majority somewhat greater than 50% for assuring the acceptance of the project. The reason is that always a certain number of Members of Parliament are not present or do not follow the party doctrine. Moreover, often the two chambers of a bicameral Parliament are elected by a different procedure[7].

Of course, the risk of political blockage is the downside of very close elections. However, in most countries a slight majority does not permit to completely modify the political regime by changing the constitution, but the government nevertheless has the power to pursuit regular policies in an efficient way[8].

In a nutshell, a relatively powerful opposition will hinder the implementation of dictatorial policies and fundamental changes in the Constitution by a "bad" politician, whereas a weak opposition is likely to result in a dictatorship as far as a "bad" party is elected to power.

Another important issue to address is that incompetent leaders are self-reinforcing. They do not only attack democracy by attempting to change the rules of the game (i.e. altering the Constitution) or by putting undemocratic policies into place. Authoritarian politicians also harm the long-run potential of democracy by indoctrinating rather than educating people. Even if a particular dictator has been overthrown in a future period, the lack of information caused by indoctrination and misinformation make it more likely that voters will also elect an incompetent politician in the future.

4 Conclusion

The present contribution has assessed how information affects the likelihood of a "bad" party being elected and its margin of victory. Informed voters have been shown to base their decision on both their preferences for redistribution and the observed level of competence of the two parties. Most of the informed voters support the more competent party. Uninformed voters cannot observe the level

[7] In Germany, for instance, the *Bundestag* (big chamber) represents the population, while the *Bundesrat* (small chamber) represents the different *Bundesländer*, which are unequal in size.

[8] A country in which one often observes small majorities, but where the President still has considerable powers, are the United States. Even though the President has the political means to enforce regular policies, the strong opposition has the opportunity to "check and balance" many key decisions of the government.

of competence of the candidates, but receive a noisy signal, which is the state of the economy. Competent incumbents are more likely to be associated with an economic boom than incompetent ones. In most states of the world the majority of uninformed voters support the more competent candidate, although with positive probability most uninformed voters support the "wrong" candidate. This is the case when the economy is booming despite the incumbent being less competent than the challenger or when the economy is in a recession despite the incumbent being more competent than the challenger. In these situations the "wrong" candidate gets elected if the proportion of informed voters is smaller than a critical level q^*. The more the proportion of informed voters is further decreased, the greater is the margin of victory of the "bad" candidate. Historical evidence suggests that massive popular support for a "bad" leader can result in authoritarian policies and a fatal threat to democracy.

The present contribution has succeeded in modelling the role played by information in elections and has provided a theory of the occurrence of undesirable electoral outcomes. The main implications of the model are consistent with empirical findings and the historical evidence. However, further research in this field is strongly encouraged. Empirical studies relating the level of information and education of the population to the stability and persistence of democracy could test the findings of the present theoretical contribution and inspire further theoretical research.

References

[1] Acemoglu, Daron and James Robinson, 2001, "A Theory of Political Transitions", *American Economic Review* 91: 938-63.

[2] Acemoglu, Daron and James Robinson, 2006, *Economic Origins of Dictatorship and Democracy*, Cambridge, Cambridge University Press.

[3] Austen-Smith, David and Jeffrey Banks, 1996, "Information Aggregation, Rationality and the Condorcet Jury Theorem", *American Political Science Review* 90: 34-45.

[4] Brender, Adi and Allan Drazen, 2005, "How Do Budget Deficits and Economic Growth Affect Reelection Prospects? Evidence from a Large Cross-Section of Countries", NBER Working Paper 11862.

[5] Caselli, Francesco and Massimo Morelli, 2004, "Bad politicians", *Journal of Public Economics* 88: 759-82.

[6] Fair, Ray, 1978, "The Effect of Economic Events on Votes for President", *Review of Economics and Statistics* 60: 159-72.

[7] Feddersen, Timothy and Wolfgang Pesendorfer, 1996, "The Swing Voter's Curse", *American Economic Review* 86: 408-24.

[8] Feddersen, Timothy and Wolfgang Pesendorfer, 1997, "Voting Behavior and Information Aggregation in Elections with Private Information", *Econometrica* 65: 1029-58.

[9] Fischer, Alastair J., 1999, "The probability of being decisive", *Public Choice* 101: 267-83.

[10] Gelman, Andrew, Gary King and W. John Boscardin, 1998, "Estimating the probability of events that have never occurred: When is your vote decisive?", *Journal of the American Statistical Association* 93: 1-9.

[11] Gul, Faruk, and Wolfgang Pesendorfer, 2006, "Partisan Politics and Aggregation Failure with Ignorant Voters", mimeo, Princeton University.

[12] Gylfason, Thorvaldur, 2001, "Natural resources, education, and economic development", *European Economic Review* 45: 847-59.

[13] Heath, Anthony, Robert Andersen and Richard Sinnott, 2002, "Do Less Informed Voters Make Mistakes? Political Knowledge and Electoral Choice", CREST working paper 97, University of Oxford.

[14] Lijphart, Arend, 1999, *Patterns of Democracy*, New Haven and London, Yale University Press.

[15] Margolis, Howard, 2001, "Pivotal voting", *Journal of Theoretical Politics* 13: 111-16.

[16] Margolis, Howard, 2002, "Pivotal voting and the emperor's new clothes", *Social Choice and Welfare* 19: 95-111.

[17] Martinelli, César, 2006, "Would rational voters acquire costly information?", *Journal of Economic Theory* 129: 225-51.

[18] Michels, Roberto, 1962 [1911], *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracies*, New York, Free Press.

[19] Moore, Barrington, 1993, *Social Origins of Dictatorship and Democracy*, Boston, Beacon Press.

[20] Mulligan, Casey and Charles Hunter, 2003, "The empirical frequency of a pivotal vote", *Public Choice* 116: 31-54.

[21] Myatt, David, 2006, "On the Theory of Strategic Voting", forthcoming in *Review of Economic Studies*.

[22] Olson, Mancur, 1993, "Dictatorship, Democracy, and Development", *American Political Science Review* 87: 567-76.

[23] Psacharopoulos, George, 1994, "Returns to investment in education: A global update", *World Development* 22: 1325-43.

[24] Wintrobe, Ronald, 1998, *The political economy of dictatorship*, Cambridge, Cambridge University Press.

Appendix A: Equilibrium with Voters Caring About Being Pivotal and with a Greater Weight Attributed to the Winning Party

In the present appendix it is first examined if under pure pivotal voting a Bayesian Nash Equilibrium with some uninformed voter j being pivotal exists. Pure pivotal voting corresponds to a situation when under uncertainty uninformed voters select the action that is optimal conditional on being pivotal. Informed voters face no uncertainty about the competence level of the incumbent and therefore choose the same actions as derived in section 2.2.1.

Consider the case of Y^+ being observed. With probability $\frac{p}{2}$ the state of the world is $E = -e$. As before, the share of informed voters voting for the challenger equals $(v_C^i \mid E = -e) = \frac{1-e}{2} < \frac{1}{2}$. Also the vote share of the uninformed voters for the challenger is below one half: $(v_C^u \mid Y^+) = \frac{1-(p-\frac{1}{2})e}{2} < \frac{1}{2}$. For any value of q the incumbent will be re-elected. Thus, an uninformed voter knows that he can never be pivotal if the state of the world is $Y^+, E = -e$. Observing Y^+ he should assign a probability 0 to being pivotal when $E = -e$.

Further, when Y^+ is observed, the probability of $E = 0$ is $\frac{1}{2}$. The votes shares of the informed and uninformed voters equal $(v_C^i \mid E = 0) = \frac{1}{2}$, respectively $(v_C^u \mid Y^+) = \frac{1-(p-\frac{1}{2})e}{2} < \frac{1}{2}$. For any value of $q < 1$ the incumbent will be re-elected. Again, an uninformed voter observing Y^+ should assign probability 0 to being pivotal when $E = 0$.

Moreover, given that Y^+ is observed, with probability $\frac{1-p}{2}$ the state of the world is $E = e$. In this case the share of informed voters supporting the challenger corresponds to $(v_C^i \mid E = e) = \frac{1+e}{2} > \frac{1}{2}$. If all uninformed voters but j play non-strategically, we have, as before, $(v_C^u \mid Y^+) = \frac{1-(p-\frac{1}{2})e}{2} < \frac{1}{2}$. For one particular level of q the total vote share obtained by the challenger does exactly equal $\frac{1}{2}$, and an uninformed voter j would be pivotal[9]. Thus, observing Y^+, an uninformed voter should assign the probability of $(E = e \mid Y^+, \textit{being pivotal}) = 1$ and vote for the challenger if $S + e > 0$. However, this result is not a Bayesian Nash Equilibrium as it requires the assumption that only player j plays strategically and all other uninformed voters play non-strategically, which would not be their best reply. But if all other uninformed voters played strategically as well, the proportion $\frac{1+e}{2} > \frac{1}{2}$ of them would vote for the challenger. This, however, implies that player j could not be pivotal anymore for $E = e$ (as the challenger would always win the election) and that his beliefs are not consistent anymore.

[9] Put more precisely, voter j could also be pivotal if one party leads with only one vote. This will be taken into account in the formal treatment further below. For the informal discussion that immediately follows we will for simplicity only refer as pivotal to the case where each party receives exactly half of the votes.

There is no level of $E = \{-e, 0, e\}$, for which an uninformed player j could ever be pivotal if all uninformed voters play strategically.

To sum up, in none of the cases discussed above a given uninformed voter j could be pivotal, as long as all uninformed voters behave strategically. Thus, there cannot exist a Bayesian Nash Equilibrium with pure pivotal voting of all uninformed voters and with some uninformed voter j being pivotal.

Above the case of Y^+ has been treated. The reasoning for Y^- is exactly the same, and also in that case under no circumstances an uninformed player can be pivotal if all uninformed players behave strategically. For this reason, the case of Y^- will not be treated in what follows.

In section 2 of the main text it has been discussed why it is realistic to assume that voters do not exclusively care about the tiny probability of being pivotal, but do at least to a small extent take the vote share of the losing party and the margin of victory into account. Below, the case of voters caring mostly, but not exclusively, about the winner and about being pivotal is discussed for a state of the world Y^+. The reasoning for Y^- would be exactly the same.

Given Y^+, the probability of $E = -e$ is $\frac{p}{2}$. As shown above, for this case the incumbent wins and j can never be pivotal. The payoffs of the uninformed voter j for voting for the challenger or the incumbent are displayed below.

$$U_j^C = n\left[\theta v_G \pi_G + (1-\theta) v_C \pi_C\right] + (1-\theta)\left[S - e\right] \tag{A1}$$

$$U_j^G = n\left[\theta v_G \pi_G + (1-\theta) v_C \pi_C\right] \tag{A2}$$

where n=number voters excluding j, θ=utililty weight given to the winning party, superscript C (G) refers to the challenger (incumbent).

With probability $\frac{1}{2}$ the state of the world is $E = 0$. As shown before, the incumbent wins for this setting and an uninformed voter j has the following payoffs.

$$U_j^C = n\left[\theta v_G \pi_G + (1-\theta) v_C \pi_C\right] + (1-\theta) S \tag{A3}$$

$$U_j^G = n\left[\theta v_G \pi_G + (1-\theta) v_C \pi_C\right] \tag{A4}$$

Further, the probability of $E = e$ is $\frac{1-p}{2}$. For this case, who wins depends on q. As shown in equation (20), the threshold level of q is labelled $q^* = \frac{p-\frac{1}{2}}{p+\frac{1}{2}}$.

If $q > q^*$ the challenger is elected, while for $q < q^*$ the incumbent is re-elected. If $q = q^*$ holds, both parties receive the same number of votes, and an uninformed voter j is pivotal[10]. If the value of q is common knowledge, an uninformed voter j would know if he can possibly be pivotal or not. Voter j could only be pivotal for the case of $q = q^*$, while for all other values of q he would

[10] As mentioned earlier, voter j can also be pivotal when there is a difference of one vote between the two parties. This will be accounted for in the formal analysis.

know that he cannot possibly be pivotal and would not include the possibility of being pivotal in the maximisation decision.

If voter j has incomplete information about the level q, but knows the distribution of possible values of q, he will take into account for the maximisation decision that with some probability he will be pivotal. This case is the most interesting and will be emphasised for the remainder of Appendix A. The probability that $q > q^*$ (whatever j plays) is denoted χ. The probability that $q = q^*$ is labelled φ. Voter j can be pivotal in three possible ways. With probability φ^- G would get elected if voter j wouldn't vote, and him voting for C would lead to a tie, where both parties get elected with equal probability. With probability φ^0 there would be a tie without voter j and j's decision decides which party wins. With probability φ^+ C has a small majority before j's vote and j's vote for G would lead to a tie. We denote $\varphi = \varphi^- + \varphi^0 + \varphi^+$.

For $E = e$ the payoffs of voting for C or G are displayed below.

$$\begin{aligned} U_j^C &= (\chi + \frac{1}{2}\varphi^- + \varphi^0 + \varphi^+)n\theta\,[S+e] \\ &\quad +(1-\chi-\frac{1}{2}\varphi^- - \varphi^0 - \varphi^+)n(1-\theta)\,[S+e] \\ &\quad +(\chi + \frac{1}{2}\varphi^- + \varphi^0 + \varphi^+)\theta\,[S+e] \\ &\quad +(1-\chi-\frac{1}{2}\varphi^- - \varphi^0 - \varphi^+)(1-\theta)\,[S+e] \end{aligned} \tag{A5}$$

$$U_j^G = (\chi + \frac{1}{2}\varphi^+)n\theta\,[S+e] + (1-\chi-\frac{1}{2}\varphi^+)n(1-\theta)\,[S+e] \tag{A6}$$

Voter j's decision has two effects. First, it affects the probability of either C or G winning, which matters for the weight given to $\pi_C = S + e$ (as before we normalise $\pi_G = 0$). Second, it provides one additional vote for a given party. This is reflected by the second line of equation (A5) (in equation (A6) this second effect is zero as $\pi_G = 0$).

Putting together the equations (A1) to (A6), the difference in utility between U_j^C and U_j^G is obtained.

$$\begin{aligned} U_j^C - U_j^G &= \frac{p}{2}(1-\theta)\,[S-e] + \frac{1}{2}(1-\theta)S \\ &\quad +\frac{1-p}{2}(\frac{1}{2}\varphi^- + \varphi^0 + \frac{1}{2}\varphi^+)n(2\theta-1)\,[S+e] \\ &\quad +\frac{1-p}{2}(1-\theta+(\chi+\frac{1}{2}\varphi^- + \varphi^0 + \varphi^+)(2\theta-1))\,[S+e] \end{aligned} \tag{A7}$$

In the main text it was shown that for Y^+ the majority of uninformed voters vote for the incumbent, although with positive probability the challenger is more competent. Below it will be assessed whether this result still holds once pivotal voting is taken into account and when the winning party receives a greater

weight in the utility function. For this analysis we shall focus on the "median" uninformed voter, i.e. set S=0. Equation (A7) becomes:

$$\begin{aligned} U_j^C - U_j^G &= -\frac{p}{2}(1-\theta)e + \frac{1-p}{2}(\frac{1}{2}\varphi^- + \varphi^0 + \frac{1}{2}\varphi^+)n(2\theta-1)e \\ &\quad + \frac{1-p}{2}\left\{(1-\theta) + (\chi + \frac{1}{2}\varphi^- + \varphi^0 + \varphi^+)(2\theta-1)\right\}e \quad \text{(A7')} \end{aligned}$$

As displayed in equation (A8) the "median" uninformed voter has greater incentives to vote for C for a higher weight given to the winning party, θ.

$$\begin{aligned} \frac{\partial(U_j^C - U_j^G)}{\partial\theta} &= (p-\frac{1}{2}) + n(1-p)(\frac{1}{2}\varphi^- + \varphi^0 + \frac{1}{2}\varphi^+) \\ &\quad + (1-p)(\chi + \frac{1}{2}\varphi^- + \varphi^0 + \varphi^+) > 0 \quad \text{(A8)} \end{aligned}$$

Similarly, we have $\frac{\partial(U_j^C - U_j^G)}{\partial\varphi^-} > 0$, $\frac{\partial(U_j^C - U_j^G)}{\partial\varphi^0} > 0$ and $\frac{\partial(U_j^C - U_j^G)}{\partial\varphi^+} > 0$, i.e. voter j is more likely to vote for the challenger the higher the probability is of being pivotal. Also $\frac{\partial(U_j^C - U_j^G)}{\partial p} < 0$, which implies that the vote share of the challenger decreases in the precision of the signal Y.

In the main text we obtained the result that for a state of the world of Y^+ the majority of the uninformed voters support the incumbent although with positive probability the challenger is more competent. This result remains valid in a framework of voters having a positive probability of being pivotal and of having a higher utility weight of the winner (i.e. $\theta > \frac{1}{2}$), as long as the probability of being pivotal, φ, and the weight bias towards the winning party, $(\theta - \frac{1}{2})$, are not too large, and the signal is precise (high p). In this case $(U_j^C - U_j^G)$ displayed in equation (A7') becomes negative and the incumbent is re-elected. Also all other results of the main text hold as long as this is the case.

Appendix B: Endogenous Determination of the Information Level of the Voters

In this appendix a simple model is built[11] that predicts under what conditions people have incentives to invest in education, knowledge and critical thought, rather than to choose ignorance. Uninformed voters will emerge when it does not pay off for people to collect information which makes autonomous decision making feasible.

A representative agent faces the time constraint displayed in equation (B1). She has to spend a part of her total time (normalised to 1) assuring her subsistence consumption s and has some surplus (1-s) which she can either use for manual labour L or invest in education and knowledge E.

[11] The notation is distinct for this appendix. The parameters used have another meaning than in the main text of the paper.

$$(1 - s) = L + E \tag{B1}$$

As shown in equation (B2), her utility u depends on the social status z, on the economic output y and on the governmental policy program v^G. The government policy is taken as exogenous in the decision process of the agent, as her impact on election outcomes is close to zero.

$$u = \theta z + y + v^G \tag{B2}$$

where θ=relative weight of the social status in the utility function.

The social status depends on both the level of education E and on other factors m.

$$z = \rho E + m \tag{B3}$$

where ρ=positive parameter.

Equation (B4) displays the production function of the economic output y.

$$y = \alpha L + qt\beta E^{\gamma} \tag{B4}$$

where α, β, γ=positive parameters (with $0<\gamma<1$, and $\alpha < \beta$), q=probability of an agreement conditional on matching with a potential trade partner, t= probability of matching with a potential trade partner.

The economic output y of an agent is composed by the (linear) returns on the fraction L that is devoted to some simple manual tasks and the fraction E that is invested in education and knowledge. The commodities produced with the input of unskilled labour L are non-sophisticated and can be consumed directly by the agent without being traded first. Alternatively, the agent can accumulate education and knowledge E which can eventually be used as skilled labour for the production of more sophisticated goods. The returns for a given unit of education E transformed in skilled labour are typically higher than the returns for a unit of unskilled labour. This is reflected by $\beta > \alpha$.

The marginal returns to education are assumed to be positive but decreasing, which is reflected by $0<\gamma<1$. This assumption is consistent with the empirical evidence on returns to schooling and education (cf. for example Psacharopoulos, 1994; Gylfason, 2001).

Further, the sophisticated commodities produced by the mean of investment in education cannot be consumed directly as they involve specialisation due to specific know-how. Therefore, a trade partner has to be found for exchanging these sophisticated goods against other commodities. The probability of successfully reaching a trade agreement is given by the probability t of meeting a trade partner times the probability q of reaching an agreement with this potential trade partner.

Introducing the equations (B1), (B3) and (B4), the utility function can be written in the following way:

$$u = \theta(\rho E + m) + \alpha(1 - s - E) + qt\beta E^{\gamma} + v^{G} \tag{B5}$$

From the first order condition of the maximisation of u with respect to E we obtain equation (B6), which displays the optimal choice of education E^*. The second order condition holds.

$$E^* = \left(\frac{qt\beta\gamma}{\alpha - \theta\rho}\right)^{\frac{1}{1-\gamma}} \tag{B6}$$

The optimal level of education is increasing in the parameters β and ρ associated to the efficiency of production of sophisticated skill-intensive goods and to the impact of education on the social status. Further, E^* increases for a higher probability of a successful trade agreement for sophisticated goods qt.

Religion, culture, geographic and economic factors matter. First, a religion and social norms that value education and knowledge would be reflected by a higher ρ, as the conformity with religious and cultural norms affects the social status. Examples of religions which traditionally have highly valued education, knowledge and critical thought include Protestantism or Judaism. Martin Luther's initial reformation movement has emphasised the importance of believers studying the bible on their own in their mother tongue (whereas before, only the thin Latin speaking elite was able to read the bible). Protestant thinkers have also stressed the importance of individual belief (as opposed to following ecclesiastic norms) for achieving salvation. Similarly, Judaism has a strong tradition of incentivising people to obtain knowledge.

Further, geographic factors matter. In a rural area the probability t of matching a suitable trade partner for sophisticated goods is much smaller than in an urban setting. Thus, people living in small villages have lower incentives to invest time in education and information.

Moreover, in a capitalist economy with social mobility and free (internal) trade it is more likely to be able to reach an agreement with a potential trade partner. If a trade agreement is mutually beneficial, it will take place in such an economy - independently of class, caste or ethnicity of the trade partners. This corresponds to a high q, and eventually a high E^*.