Imperfect Information, Democracy, and Populism*

Johannes Binswanger†

Jens Prüfer‡

November 17, 2008

Abstract

The modern world is complex and difficult to understand for voters, who may hold beliefs that are at variance with reality. Politicians face incentives to pander to voters' beliefs to get reelected. We analyze the welfare effects of this pandering and show that it comes along with both costs and benefits. Moreover, we explore optimal constitutional design in the presence of imperfect information about how the world works. We compare indirect democracy to direct democracy and to delegation of policy making to independent agents. We find that indirect democracy is often welfare optimal.

Key words: Accountability, beliefs, imperfect knowledge, democracy, experts, populism, politics, voting.

JEL classification: D72, D78, D83.

*We would like to thank Jan Boone, Reto Föllmi, Alia Gizatulina, Manuel Oechslin, Jan Potters, and seminar participants at various universities for very helpful comments. All errors are our own.

†Department of Economics, CentER, Netspar, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; j.binswanger@uvt.nl.

‡Department of Economics, CentER, Tilec, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; j.prufer@uvt.nl.

1 Introduction

The best argument against democracy is a
five-minute conversation with the average voter.

Winston Churchill

The world is complex and people may not properly understand how the modern society and its economy function. In some cases, voters' beliefs deviate substantially from the view of experts. Caplan (2001) conducted a survey where he compares the opinions of the general public to the opinions of economists with regard to a number of economic issues. For instance, he asked people whether they would think that trade agreements between the United States and other countries have helped create more jobs in the U.S. He coded the answer that trade agreements "have cost the U.S. jobs" as 0, that they "haven't made much of a difference" as 1, and that they "helped create jobs in the U.S." as 2. He finds that the mean response among economists is 1.47. In contrast, the mean response among the general audience is only .64. Caplan reports similar discrepancies with respect to other economic issues.

As a result of imperfect knowledge about a complex world, voters may not always be able to judge what policies are truly in their best interest and, as a result, hold incorrect beliefs. Politicians who aim to get reelected have an incentive to pander to voters' beliefs and hence to potentially distort policies.

In this paper, we investigate the welfare consequences of politicians' pandering to voters' beliefs in the case of indirect democracy. Second, we explore under what conditions other forms of government may lead to higher welfare. In particular, we compare indirect democracy to direct democracy and to the case where policy making is delegated to independent agents that are experts in a particular field.

In our model, welfare depends on a unidimensional policy action. In the case of indirect democracy, policy is set by an office-holding politician. In the case of direct democracy, it is determined by voters themselves whereas in the case of independent agents it is set by the latter.

We capture voters' imperfect information by assuming that no voter observes which level of the policy action is truly in the best interest of themselves and society. Rather, voters hold prior *beliefs* about the welfare effects of policy actions.

In the case of indirect democracy, there are two types of politicians in the model. The first type is dubbed competent. This means that she obtains a perfectly revealing signal about the policy action that is best for society from an ex ante point of view. In contrast, the incompetent politician receives a noisy signal about the optimal policy and is thus imperfectly informed about the world. The assumption of the existence of incompetent politicians refers to the fact that politicians may have incompetent advisors, or they may have ideological views (as may voters). Politicians' prime objective is to get reelected.

Voters aim to reelect a politician that carries out policies that appear welfare maximizing looking at the world through the lenses of their beliefs. They anticipate that such a politician would also lead to the highest expected welfare in the next office period, given their beliefs. This characteristic induces an incentive for politicians to cater for the pivotal voter's beliefs in order to maximize their chance of getting reelected. Thus, populist policies arise endogenously.

As it is usual in dynamic games of incomplete information, voters' need to have a belief about the behavior of the two politician types in order to be able to update their beliefs about an incumbent politician's type. Our analysis is based on a type of beliefs which includes an element of bounded rationality but contains full rationality as a limit case. These beliefs are related to the cognitive hierarchy model by Camerer et al. (2004). In a nutshell, we assume that voters may anticipate a politician's strategic response to their beliefs and, as a result, adapt their beliefs by one additional order not more than k times. This gives rise to what we will dub sophistication-k beliefs. The number k may be finite or, in the case of unlimited rationality, infinite.

So far, there has been little attempt to incorporate elements of behavioral economics and bounded rationality into models of voting and this seems to be an important item on the research agenda in political economy. For instance, Besley (2006) writes "going forward it would be interesting to understand better what the differences are between behavioral models of politics and the postulates of strict rationality [...]" By basing our analysis on sophistication-k beliefs, we make one step in this direction.

Our analysis shows that, under indirect democracy, the policy action is determined as a weighted average of a politician's signal and the pivotal voter's prior beliefs. Thus, policy making is partially populist. Importantly, this populism comes along with both costs and benefits. The costs refer to the fact that the competent politician has an incentive to partially ignore her signal which means wasting useful information. On the other hand, populism also induces the incompetent politician to partially ignore her noisy signal which may be beneficial. We then compare indirect democracy to direct democracy and to delegation of policy making to non-accountable agents (i.e. experts).

Due to the "weighted-average" nature of policies under indirect democracy, we find that it can be expected to lead to the highest welfare under many circumstances. One important exception is the case where policy making mainly requires application of formal technical knowledge. In this case, delegation of policy making to non-accountable experts is found to be optimal. Another exception is the case where judging the effects of a policy mainly requires informal knowledge which is not readily available to politicians or experts. In this case, direct democracy is found to be optimal. Overall, our analysis may help explain why indirect democracy is so prevalent around the world.

Our analysis is related to a number of existing studies, most notably Maskin and Tirole (2004). These authors also analyze the optimality of the three institutions direct democracy, indirect democracy and independent agents. They consider a binary policy choice where one policy is more popular among voters than the other. Furthermore, they assume that politicians are intrinsically motivated to carry out certain policies. Our analysis differs in two main ways. Technically, we allow policies to lie on the entire real line. This allows us to conceptualize in a natural way notions of imperfect knowledge about the world, such as the distance of voters' beliefs from the truth or the noisiness of a politician's signal. Second, key factors in our analysis are the reliability of politicians' and non-accountable agents' information, which do not appear in the analysis of Maskin and Tirole. We view our work as complementary to theirs. A study that is related to Maskin and Tirole (2004) is Canes-Wrone et al. (2001). These authors analyze the incentive of politicians' to pander to voters' beliefs in a binary setup that differs substantially from ours.

Alesina and Tabellini (2007, 2008) provide an in-depth analysis of the advantages and disadvantages of accountability. In particular, they compare the performance of a politician who aims to get reelected with the performance of a bureaucrat who is concerned about her career perspective. Populism and imperfect knowledge about the world do not play an essential role in their analysis. Schultz (2008) analyzes the welfare effects of accountability by focusing on the term length of office periods. In our study, we take this term length as given.

The rest of this paper is organized as follows. Section 2 introduces our model of indirect democracy. In Section 3 we solve the model. In Section 4 we characterize welfare under indirect democracy. In Section 5 we compare indirect democracy to the case of direct democracy and to delegation of policy making to non-accountable agents. In Section 6 we discuss our findings. Finally, in Section 7 we conclude.

2 A Model of Indirect Democracy

Voters

We consider an economy populated by a unit mass of individuals to which we refer as voters. Voters have an $identical$ utility function given by

$$V = -\left(g - x^{*} - \varepsilon\right)^{2}. \tag{1}$$

This utility function applies to a given office period. We defer a discussion of timing until later. The variable g is unidimensional and denotes a policy action. Under indirect democracy, g is set by the office-holding politician. Neglecting ε, the utility maximizing level of g is given by

x^*, which may lie on the entire real line. The crucial assumption in our framework is that x^* is unobserved. The variable ε is a normally distributed random variable with an expected value of zero and a variance of σ_ε^2. As will be discussed in more detail later, we assume that, within an office period, nature first draws x^*, before ε is realized. Thus, ε represents an "interim" shock to x^*. As is the case with x^*, ε is also unobserved.

To consider an example, suppose that there is a given budget to be spent for combating crime. Suppose that the relevant decision is to determine the share of this budget to spend on preventive measures (schooling, prevention of youth unemployment, quality of neighborhoods etc.) versus the share to spend on punishment (e.g. prison infrastructures). In this example, x^* would refer to the optimal budget share for preventive measures, given the $general$ current situation in society. This may refer to the degree of inequality and ethnic heterogeneity, the degree to which people follow certain norms or cultural practices, the general level of youth unemployment etc. The variable ε would then correspond to a shock to the "threat of crime" and may originate from an sudden rise in youth unemployment, a sudden increase in immigration or the like.

An alternative example for x^* may be the ex ante optimal number and types of admissible investment funds that people may invest their retirement savings within a defined contribution system. In this case, the shock ε would correspond to short-term random factors such as the development of the stock market, the attention paid to pension issues by the press etc. To make an extreme example, if ex post the returns of a particular company stock has been unusually high, it would have been optimal to offer a retirement preparation fund that would have invested only in this company. From an ex ante point of view, this would unlikely have been optimal.

The quadratic specification (1) in form of a "loss function" is chosen for tractability. This utility function should be taken as reflecting indirect utility, meaning that optimal values of all other choices that voters may make are already substituted. The essential features of (1) are the following two. First, x^* determines a unique interior optimum for g from an ex ante point of view, i.e. before ε is realized.[1] Ex post, the welfare maximizing level of g is $x^* + \varepsilon$. Second, there is risk aversion over the realizations of g.

As already stated, x^* is not observed. However, voters have prior beliefs about x^*. Specifically, we make the following assumption.

Assumption 1 *Each voter i assumes that $V^i = -(g - x^i - \varepsilon)^2$ and has a prior belief that x^i is normally distributed with mean μ^i and variance σ_x^2.*

Assumption 1 implies that the prior means of x^i may be heterogeneous among voters while, for simplicity, we assume that the variance is common across voters.[2] We also assume that the distribution of ε is common knowledge among voters.[3]

Politicians

Under indirect democracy, the policy action g is chosen and implemented by an incumbent politician. An incumbent politician's objective in the first office period is to get reelected for a second term. As discussed below, we analyze a model with two office periods and we assume that in the second period politicians simply maximize welfare. The latter assumption is to be understood as a shortcut and does not affect our main conclusions in a substantive way. [4]

The politician chooses g without observing ε. A politician is assumed to know the distribution of μ^i. Moreover, we assume that politicians are imperfectly informed and do not a priori know x^*, too. Rather, they receive a signal ξ that is informative about the value of x^*.

[1] If the utility function were linear, then the optimal level of g would lie at the boundary of an admissible region for g, except in the case where voters were indifferent between any level of g

[2] Applying the median voter theorem will require restricting heterogeneity on μ^i later on.

[3] We use the normal distribution because of the high tractability of its updating formulas (see below). For the example of choosing a share of a budget to spend on preventive measures for combating crime, the policy variable could only take on values between zero and one and would not be consistent with a normal distribution. However, it is straightforward to find a transformation of the domain of admissible policies such that they may take on any real value. For instance, any function that is bijective and maps $[0, 1]$ on the entire real line would achieve this for budget shares, e.g. an appropriately shifted tangent function.

[4] In particular, we may allow for rent seeking in the second period along the lines of a model discussed in Persson and Tabellini (2000), Ch. 4. See footnote 10. We exclude rent seeking here in the interest of transparency of the analysis.

We assume that there are two types of politicians that we dub competent and incompetent, respectively. The prior probability that a politician is competent is denoted by α and is common knowledge. In case of the competent politician, $\xi = x^*$, i.e. the signal reveals the truth. An incompetent politician receives a noisy signal $\xi = x^* + \zeta$, where ζ is a random variable with mean zero and variance σ_ζ^2.[5] Importantly, a politician does not observe her type.

We assume that the distribution of ζ is common knowledge. In contrast, the overall distribution of ξ cannot be common knowledge since x^* is not commonly observed. If it were, then the incompetent politician type could infer x^* and the existence of the incompetent type would not be substantive.

Specifically, we make the following assumption about politician's beliefs about x^*.

Assumption 2 *Each politician believes that $E[x^* \mid \xi] = \xi$.*

As we have done in the the case of voters, we could have made a more explicit assumption that politicians have a prior believe that x^* is normally distributed with expected value ξ. The consequences of this more specific assumption are minor and we discuss it briefly in Appendix B. In the main analysis, we assume that a politician bases her behavior only on $E[x^* \mid \xi]$.

In principle, it may be natural to allow for $E\zeta \neq 0$. In particular, one may argue that politicians are drawn from the general population and may thus have systematically biased views about x^*. We briefly discuss this case in Appendix C but do not consider it in the main part of the analysis since it complicates the analysis without leading to substantive additional insights.[6]

In reality, several factors may be decisive for a politician's competence. An important factor is the availability of a competent staff of advisors. Depending on institutions, these advisors may be mainly chosen by a politician's party or the politician may select them herself. In the latter

[5] The assumption that the competent politician perfectly observes x^* is made for convenience. The main conclusions from our analysis could also be obtained if the competent politician received a noisy signal where the variance of the noise term is smaller than for the incompetent politician.

[6] Whether politicians should be understood as a "representative sample" drawn from the general population clearly depends on the nature of the political selection process. While it may be representative in some countries, it is more elitist and hence biased in others.

case, the ability to screen out good advisors is also key for a politician's effective competence. Second, a politician needs to have the ability to listen to other views and to change her view in the presence of compelling rational reasons to do so. Third, a politician's competence may also be affected by the degree to which she is ideological. All these factors relate to our model in that they determine the magnitude of σ_ζ^2.

The Political Game

We are now in a position to discuss the timing of events. Following Maskin and Tirole (2004) and Schultz (2008), we consider two subsequent office periods, as already mentioned. Considering a two-period model provides a natural first step to understanding the issues at stake and may lie the ground for the analysis of more complex setups.

The timing of events is as follows. At the beginning of the first office period, nature selects x^* for the first period, which we henceforth denote by x_1^*. Nature also draws an incumbent politician and determines his signal ξ_1, i.e. $\xi_1 = x_1^*$ for the competent and $\xi_1 = x_1^* + \zeta_1$ for the incompetent politician. In the next stage, the incumbent politician chooses a level of g_1, i.e. the policy action for the first office period. Importantly, she does so without observing ε_1. After g_1 has been set, ε_1 is realized, hence V_1. Voters observe the sum $x_1^* + \varepsilon_1$ and update their beliefs about x_1^* and about the incumbent politician's type. Then, the election takes place. The incumbent is reelected if she gains at least half of the votes. Otherwise she is ousted and replaced by a challenger of unknown type.

Then, the second office period starts. Nature draws a new x^*, denoted by x_2^*. We allow for the special case where $x_2^* \equiv x_1^*$ but do not need to require any restrictions on x_2^*. If the politician in office is a reelected incumbent that is competent, his signal is $\xi_2 = x_2^*$. In case of a reelected incumbent that is incompetent, the signal is $\xi_2 = x_2^* + \zeta_1$, i.e. she keeps her realization of ζ_1. Thus the deviation of her signal from the "truth" x_t^* remains constant over time. In case of a challenger, we have $\xi_2 = x_2^*$ for the competent type and $\xi_2 = x_2^* + \zeta_2$ for the incompetent type. We assume that the distribution of ζ remains constant over time. The politician in office chooses

g_2 without observing ε_2. Finally, nature draws ε_2 and V_2 is realized. The distribution of ε also remains identical over time.

3 Analysis of Indirect Democracy

3.1 The Reelection Decision

In order to prevent confusion, we start with a remark on notation. For any random variable z, we write z_t^i for a voter i's *prior* belief about the distribution of z in office period $t = 1, 2$. A hat is used for denoting posterior beliefs. Thus, $\hat{z}_t^i$ refers to a voter i's posterior belief about the distribution of z in period t. We apply the same convention for parameters of random variables. If the distribution of a random variable remains constant over time (ε and ζ), we sometimes omit the time subscript for parameters. Except in special cases, we will not make a notational distinction between random variables and their realizations. As a last remark on notation, we will often say that a voter i's belief about x_1^* is that it is distributed according to x_1^i or $\hat{x}_1^i$. This is a shortcut for saying that, from the perspective of voter i not observing x_1^*, the latter is substituted by a random variable about which the voter has prior or posterior beliefs that it is distributed according to x_1^i or $\hat{x}_1^i$, respectively.

Voters make their reelection decision after having observed g_1 and $x_1^* + \varepsilon_1$. This allows voters to update their beliefs about x_1^* and also about the realization of ζ_1 under the hypothesis that the incumbent politician is incompetent. Before analyzing the reelection decision, we thus characterize how voters update these beliefs.

Updating Beliefs about x_1^* and ζ_1

According to Assumption 1, voter i initially holds a prior belief x_1^i about x_1^*. The prior belief entails that x_1^i is normally distributed with mean μ_1^i and variance σ_{1x}^2. Before making his reelection decision, a voter observes $x_1^* + \varepsilon_1$.[7] This allows him to update his beliefs about x_1^*. Due to the assumption that prior beliefs are normal, we obtain a very tractable updating formula stated in the following lemma. (Proofs are always given in Appendix A if not stated otherwise.)

Lemma 1 (Posterior beliefs about $\boldsymbol{x}^*$) *Suppose that voter i's belief about x_1^* is that it is distributed according to the random variable x_1^i. Assume that the prior belief is that x_1^i is normally distributed with mean μ_1^i and variance σ_{1x}^2. Then, the posterior belief is that x_1^i is normally distributed with mean $\hat{\mu}_1^i = \mu_1^i + \beta\left(x_1^* + \varepsilon - \mu_1^i\right)$ and variance $\hat{\sigma}_{1x}^2 = \frac{\sigma_{1x}^2 \sigma_\varepsilon^2}{\sigma_{1x}^2 + \sigma_\varepsilon^2}$, where $\beta \equiv \frac{\sigma_{1x}^2}{\sigma_{1x}^2 + \sigma_\varepsilon^2}$.*

Note that the degree of updating depends on the signal-to-noise ratio $\beta = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2}$. While this result is highly standard, it is useful to explain its meaning in the context of the current analysis. Consider again the example of what share of a given budget to spend on preventive measures to combat crime. Ex ante, the optimal share is given by x_1^* and voter i believes that expected welfare is maximized by setting $g_1 = \mu_1^i$. Ex post, the voter observes V_1 (say the number of crimes) and the ex post optimal budget share for preventive measures $x_1^* + \varepsilon_1$. The latter depends on the actual threat of crime according to random short-term factors. If σ_ε^2 is very low, then observing $x_1^* + \varepsilon_1$ is very informative about x_1^* and the voter will update his beliefs strongly. If σ_ε^2 is high, the voter cannot expect to learn much and will update his beliefs only in a minor way.

In particular, suppose that a voter believes that a high share of the budget to combat crime should be spent on punishment and that the actual share spent on punishment has indeed been high. Suppose that, ex post, the crime rate is high. If σ_ε^2 were low, then the voter would infer that his prior beliefs were probably wrong. But if σ_ε^2 is high, he will conclude that criminal threat must have been unusually high.

As a special case, voters' prior beliefs may be understood as ideologies, e.g. about the desirability of capital punishment. An important determinant of ideologies is that they are

[7] Due to the symmetry of the utility function (1), a voter cannot infer $x_1^* + \varepsilon_1$ from observing g_1 and V_1 directly but only the distance between g_1 and $x_1^* + \varepsilon_1$. It is therefore convenient to assume that voters observe $x_1^* + \varepsilon_1$ directly.

persistent, i.e. people are not willing to let their ideologies "erode" (Bénabou, 2008). In light of the above discussion, we can capture ideological beliefs by assuming that σ_ε^2 is high. In this case, voters would blame ε for any observation that is at variance with their beliefs.

We now turn to ζ_1. Ex ante, voters have a common prior that ζ_1 has a mean of zero and a variance $\sigma_{1\zeta}^2$. Observing g_1 allows voters to update their beliefs about the ζ_1 associated with the incumbent politician under the hypothesis that she is incompetent.

Suppose the voters believe that the incumbent sets g_1 as a function of her signal according to $g_1 = G(\xi_1)$. The function G is undetermined at the moment since it is determined in equilibrium. The only assumption that we will impose is that it is one-to-one and onto, such that the inverse G^{-1} exists. This assumption will be verified in Subsection 3.3. Note that, in the case of the incompetent politician, g_1 must indeed be a function of ξ_1 and cannot depend on x_1^* and ζ_1 separately, since the politician does only observe the sum of both.

It follows that voter i expects that, in case of the competent politician, $G^{-1}(g_1) = x_1^*$ while, in case of the incompetent politician, $G^{-1}(g_1) = x_1^* + \zeta_1$. Thus, under the hypothesis that the incumbent politician is incompetent, and given his beliefs G, voter i can infer the sum $x_1^* + \zeta_1$ from g_1. The voter can use this information to update his beliefs about the realization of ζ_1. The updated beliefs are given in the following lemma. This lemma incorporates the fact that a rational voter would first update beliefs about x_1^* before updating beliefs about ζ_1.

Lemma 2 (Posterior beliefs about ζ_1) *Suppose that voter i's prior belief about ζ_1 is that it is normally distributed with mean 0 and variance $\sigma_{1\zeta}^2$. Furthermore, assume that the function $G^{-1}(g_1)$ exists. Then, under the hypothesis that the incumbent is incompetent, the posterior belief is that ζ_1 is normally distributed with mean $E\hat{\zeta}_1^i = \gamma\left(G^{-1}(g_1) - \hat{\mu}_1^i\right)$ and variance $\hat{\sigma}_{1\zeta}^2 = \frac{\hat{\sigma}_{1x}^2\sigma_{1\zeta}^2}{\hat{\sigma}_{1x}^2+\sigma_{1\zeta}^2}$, where $\gamma \equiv \frac{\hat{\sigma}_{1\zeta}^2}{\hat{\sigma}_{1x}^2+\sigma_{1\zeta}^2}$.*

The proof is almost identical to the one of Lemma 1 and is omitted.

The Reelection Decision

We now discuss a voter's reelection decision, given the general type of beliefs that $g_1 = G(\xi_1)$. We will make specific assumptions about beliefs in Subsection 3.3

In order to make an election decision, a voter needs to project his expected utility from reelecting the incumbent, taking into account any posterior information he has about the incumbent. He has to compare this to projected expected utility from ousting the incumbent and electing a challenger. Since a challenger is drawn from the general population of politicians, the voter can only make use of prior beliefs for projecting expected utility in the latter case. Voter i will cast his ballot for the incumbent if and only if expected utility from reelecting the incumbent is higher than expected utility from electing a challenger.

We first consider voter i's posterior probability that the incumbent is competent, given observation of g_1. This observation allows the voter to infer ξ_1 according to his beliefs G. Here, it is convenient to introduce a precise notation by denoting the realization of ξ_1 *as inferred by the voter* according to his beliefs G (or, equivalently, G^{-1}) by $\bar{\xi}_1$. Under the hypothesis that the incumbent politician is competent, $\bar{\xi}_1 = x_1^*$. More formally: $\bar{\xi}_1$ is a realization of the normally distributed random variable $\hat{x}_1^i$ from the voter's point of view who does not observe x_1^* but holds posterior beliefs $\hat{x}_1^i$ about x_1^* . Under the hypothesis that the incumbent is incompetent, $\bar{\xi}_1 = x_1^* + \zeta_1$. Thus, $\bar{\xi}_1$ is a realization of the normal random variable $\hat{x}_1^i + \hat{\zeta}_1^i$ under voter i's posterior beliefs. Using Bayes' law, the voter can determine the posterior probability that the incumbent is competent, which we denote by $\hat{\alpha}^i$.

Denote by f_c^i the density function associated with $\hat{x}_1^i$, and by f_{ic}^i the density functions associated with $\hat{x}_1^i + \hat{\zeta}_1^i$. The subscripts "c" and "ic" stand for competent and incompetent, respectively. Using this notation, the posterior probability that the incumbent politician is competent is determined as follows.

Lemma 3 (Updating $\boldsymbol{\alpha}$) *(i): Assume that the function $G^{-1}(g_1)$ exists. Then voter i's posterior probability that the incumbent is competent, conditional on observing $\bar{\xi}_1$, is given* *by*

$$\hat{\alpha}^i\left(\bar{\xi}_1\right) \equiv Pr[\text{competent}\left|\bar{\xi}_1\right.] = \frac{\alpha f_c^i\left(\bar{\xi}_1\right)}{\alpha f_c^i\left(\bar{\xi}_1\right) + (1-\alpha) f_{ic}^i\left(\bar{\xi}_1\right)}. \tag{2}$$

(ii): $\hat{\alpha}^i$ is a strictly decreasing function of $\left|\bar{\xi}_1 - \hat{\mu}_1^i\right|$.

Lemma 3.(ii) will be crucial for equilibrium characterization.

Next, we consider voter i's projection of expected utility in case of reelection of the incumbent. With probability $\hat{\alpha}^i$ the incumbent is competent. In this case, we will have $g_2 = x_2^*$. This follows from the assumption that politicians maximize expected welfare in the second period, conditional on receiving the signal ξ_2. Voter i does not observe x_2^* but substitutes his beliefs about x_2^*, denoted by x_2^i.[8] Thus, $E\left[V^i\left(g_2\right)\right] = -E\left(x_2^i - x_2^i - \varepsilon_2\right)^2 = -\sigma_\varepsilon^2$. In case that the incumbent is incompetent, we have $E\left[V^i\left(g_2\right)\right] = -E\left(x_2^i + \hat{\zeta}_1^i - x_2^i - \varepsilon_2\right)^2 = -\left[E\left(\hat{\zeta}_1^i\right)^2 + \hat{\sigma}_{1\zeta}^2 + \sigma_\varepsilon^2\right]$. Importantly, the voter uses his posterior belief about ζ_1, since the incumbent keeps his realization of ζ_1. Overall, expected utility from reelecting the incumbent is given by

$$E\left[V^i\left(g_2\right)\right] = -\sigma_\varepsilon^2 - \left(1 - \hat{\alpha}^i\right)\left[E\left(\hat{\zeta}_1^i\right)^2 + \hat{\sigma}_{1\zeta}^2\right]. \tag{3}$$

Expected utility from a challenger is determined very similarly. There $\hat{\alpha}^i$ has to be replaced by α and $\hat{\zeta}_1^i$ by $\zeta_2 = \zeta_1$. We then obtain

$$E\left[V^i\left(g_2\right)\right] = -\sigma_\varepsilon^2 - (1-\alpha)\,\sigma_{1\zeta}^2. \tag{4}$$

A voter reelects the incumbent if and only if expected utility as given by (3) exceeds expected utility as given by (4). Simple rearranging leads to the condition stated in the below lemma.

Lemma 4 (Reelection Decision) *Voter i reelects the incumbent if and only if*

$$\frac{1-\hat{\alpha}^i}{1-\alpha}\,\frac{\left(E\hat{\zeta}_1^i\right)^2 + \hat{\sigma}_{1\zeta}^2}{\sigma_{1\zeta}^2} \leq 1. \tag{5}$$

[8] We allow for arbitrary x_2^*, such that x_2^* need not depend on x_1^*. Of course, this allows for the special case where $x_2^* = x_1^*$. In this case, we would have $x_2^i = \hat{x}_1^i$.

Both $\hat{\alpha}^i$, $E\hat{\zeta}_1^i$ are functions of $\bar{\xi}_1$ (see Lemma 2 and 3).

To understand condition (5), consider first the limit case in which the voter would not learn anything about ζ_1 from inferring $\bar{\xi}_1 = G^{-1}(g_1)$, i.e. when $\hat{\zeta}_1 = \zeta_1$. In this case, $E\hat{\zeta}_1^i = 0$ and $\hat{\sigma}_{1\zeta}^2 = \sigma_{1\zeta}^2$. Thus, condition (5) simplifies to $\hat{\alpha}^i \geq \alpha$.

In the general case where voters do update their beliefs about ζ_1, reelection of the incumbent is compatible with $\hat{\alpha}^i < \alpha$ if $\left(E\hat{\zeta}_1^i\right)^2 + \hat{\sigma}_{1\zeta}^i$ is sufficiently smaller than $\sigma_{1\zeta}^2$. This means that observing $\bar{\xi}_1$ allows the voter to make a sufficiently sharp update of his beliefs about ζ_1 under the hypothesis that the incumbent is incompetent. In this case, reelecting the incumbent, the voter expects a relatively small variance associated with g_2 relative to the variance associated with g_2 when being set by a challenger. This comes at a benefit since voters are risk averse over g_2.

The posterior probability $\hat{\alpha}^i$ is a strictly decreasing function of $\left|\bar{\xi}_1 - \hat{\mu}_1^i\right|$as stated in Lemma 3. Furthermore, $\left(E\hat{\zeta}_1^i\right)^2$ is a strictly increasing function of $\left|\bar{\xi}_1 - \hat{\mu}_1^i\right|$ from Lemma 2. (Remember that $G^{-1}(g_1) = \bar{\xi}_1$.) Since $\hat{\alpha}^i$ enters negatively on the left-hand side of (5) while $\left(E\hat{\zeta}_1^i\right)^2$ enters positively, it follows that (5) holds if and only if $\left|\bar{\xi}_1 - \hat{\mu}_1^i\right|$ is sufficiently small. Using Lemma 1, we can state that voter i reelects the incumbent if and only if

$$\bar{\xi}_1 \in \left[(1-\beta)\,\mu_1^i + \beta\left(x_1^* + \varepsilon_1\right) \pm \delta^{crit}\right] \equiv I_{re}^i,$$

where δ^{crit} is a strictly positive real number that is common across voters and is a function of σ_{1x}^2, σ_{ε}^2, σ_{ζ}^2. We dub I_{re}^i voter i's reelection interval.

This observation naturally relates to the discussion of sufficient conditions for the median voter theorem to apply. Indicate the i associated with the median of μ_1^i by the superscript m. The median voter theorem holds if (1) $\bar{\xi}_1 \in I_{re}^m$ implies that $\bar{\xi}_1 \in I_{re}^i$ for at least half of voters and, conversely, (2) $\bar{\xi}_1 \notin I_{re}^m$ implies that $\bar{\xi}_1 \notin I_{re}^i$ for at least half of voters. The following lemma states two sufficient conditions for this to hold.

Lemma 5 (Median Voter Theorem) *Sufficient conditions for the median voter theo-*

rem to hold are either of the following:

(i) $\mu_1^m = \mu_1^i$ *for at least half of voters;*

(ii) $\mu_1^{\max} - \mu_1^{\min} \leq 2\delta^{crit}/\left(1-\beta\right)$*;*

where $\mu_1^{\min} \equiv \min\left\{\mu_1^i\right\}$ *and* $\mu_1^{\max} \equiv \max\left\{\mu_1^i\right\}$*.*

The meaning of the first condition is obvious. The second condition limits the range of μ_1^i. There are other sufficient conditions, but we do not explore this issue any further here. In what follows we will always assume that the distribution of μ_1^i is such that the median voter theorem applies.

3.2 Politician Behavior

In this subsection we characterize politician behavior for beliefs of the general form $g_1 = G\left(\bar{\xi}_1\right)$. In the next subsection, we consider a specific type of beliefs for which we obtain our main results.

The objective function of a politician is to maximize the probability that (5) holds for the median voter. This entails minimizing the left-hand side of (5), given the signal ξ_1. From Lemma 3, $\hat{\alpha}^m$ strictly decreases in $\left|\bar{\xi}_1 - \hat{\mu}_1^m\right|$. Furthermore, from Lemma 2, $\left(E\hat{\zeta}_1^m\right)^2$ strictly increases in $\left|\bar{\xi}_1 - \hat{\mu}_1^m\right|$ since $\bar{\xi}_1 = G^{-1}\left(g_1\right)$. It follows that a politician maximizes the probability that (5) holds for the median voter by setting g_1 such that $\bar{\xi}_1 = G^{-1}\left(g_1\right) = E\left[\hat{\mu}_1^m \left|\xi_1\right.\right]$.

Lemma 6 (Politician's Best Response) *Assume that* $G^{-1}\left(g_1\right)$ *exists. Then the incumbent politician sets* g_1 *such that* $G^{-1}\left(g_1\right) = E\left[\hat{\mu}_1^m \left|\xi_1\right.\right] = \mu_1^m + \beta\left(\xi_1 - \mu_1^m\right)$*.*

3.3 Beliefs of Sophistication of Degree *k*

The beliefs our equilibrium analysis is based upon contain an element of bounded rationality, but full rationality is obtained as a limit case. As already indicated in the label for these beliefs, they are related to the cognitive hierarchy model of Camerer et al. (2004). We develop these beliefs here in a somewhat exploratory way.

We start with what we *define* beliefs of sophistication of degree 0. These consist of believing, naively as it turns out, that an incumbent politician tries her best to maximize expected welfare, given her information ξ_1. A politician does not know ε_1 ex ante and voters are assumed to be aware of this. Therefore, beliefs of degree of sophistication k entail believing that the incumbent sets $g_1 = \xi_1$ because of Assumption 2. This applies to both, the competent and the incompetent incumbent. Relating this to the previous analysis, it follows that $G(\xi_1) = \xi_1$.

A politician's best response to these beliefs is determined by Lemma 6. It implies that $g_1 = E[\hat{\mu}_1^m | \xi_1] = \mu_1^m + \beta(\xi_1 - \mu_1^m)$. Thus, a politician's best response deviates from voters' beliefs that $g_1 = \xi_1$.

If beliefs are of sophistication of degree 1, then, by definition, a voter anticipates an incumbent politician's incentives to deviate from beliefs of sophistication of degree 0 by one order. Thus, for $k = 1$ we have by definition that $G(\xi_1) = \mu_1^m + \beta(\xi_1 - \mu_1^m)$. A politician's best response to these beliefs is again determined by Lemma 6 and we obtain $G^{-1}(g_1) = 1/\beta g_1 - (1-\beta)/\beta\mu_1^m = E[\hat{\mu}_1^m | \xi_1]$. Using Lemma 1, this implies $g_1 = \mu_1^m + \beta^2(\xi_1 - \mu_1^m) \neq G(\xi_1)$.

If beliefs are of sophistication of degree 2, then voters would also foresee this second-order incentive of a politician to deviate from their beliefs of level $k = 1$ and would expect $g_1 = \mu_1^m + \beta^2(\xi_2 - \mu_1^m)$. But then a politicians would have a third-order incentive to deviate etc. Both $G(\xi_1)$ and g_1 converge when k approaches infinity, but we do not restrict the analysis to this limit case since, in our view, it is not particularly realistic.

Our approach is to take k as exogenously given and identical across voters. We take it as a constraint on the sophistication of voters' strategic thinking. Alternatively, it may also be interpreted as a belief of voters about the strategic sophistication of the incumbent politician. The evidence discussed in Camerer et al. (2004) suggests that experimental subjects are able to foresee about one or two rounds of strategic reactions. Thinking about the chess game makes it salient how difficult it is in practice to anticipate higher-order strategic reactions of other players. We define level-k beliefs of strategic sophistication as follows.

Definition 1 (Sophistication-k beliefs) *Under beliefs of sophistication of degree k, voter i believes that $g = \mu_1^m + \beta^k (\xi_1 - \mu_1^m)$.*

For these beliefs, we simply have $G(\xi_1) = \mu_1^m + \beta^k (\xi_1 - \mu_1^m)$. Thus, G is strictly monotonic in ξ_1 if and only if k is finite. Hence the inverse G^{-1} exists if and only if k is finite. Thus, Lemma 2, 3, 6 apply for finite k but not for the limit case of an infinite k. This does not pose a problem since it is very straightforward to derive the equilibrium when k is infinite. See next subsection.

3.4 The Political Equilibrium

We start the discussion of the equilibrium with a formal definition of populism.

Definition 2 (Populism) *A politician's choice is populist if it does not only depend on her signal ξ but also on the prior belief of the median voter μ_1^m.*

Our main result that characterizes the outcomes in an indirect democracy in the first office period is the following.

Proposition 1 (Equilibrium First Period) *Suppose voters hold beliefs of degree of sophistication k. (i): If $0 \leq k < \infty$, then there exists a unique equilibrium in which the competent politician chooses*

$$g_1 = \mu_1^m + \beta^{k+1} (x_1^* - \mu_1^m), \tag{6}$$

while an incompetent politician chooses

$$g_1 = \mu_1^m + \beta^{k+1} (x_1^* + \zeta_1 - \mu_1^m). \tag{7}$$

The equilibrium always entails partial populism. (ii): If k is infinite, then there exists a pooling equilibrium where both politician types set $g_1 = \mu_1^m$, given supporting off-equilibrium beliefs. This equilibrium is perfectly populist.

Part (i) of Proposition 1 shows that g_1 is equal to a weighted average of the politician's signal about x_1^* and the median voter's prior belief about x_1^*. Remember that the signal of the competent politician is equal to x_1^* while the signal of the incompetent politician is equal to $x_1^* + \zeta_1$. The most important observation is that any equilibrium involves pandering to the median voter's beliefs and thus a populist policy choice. The degree to which policy making is populist is the higher, the lower β^k. Thus, β^k can be understood as indicating the susceptibility to populism. Since $0 < \beta < 1$, a higher k implies a higher susceptibility to populism. The intuition behind this result follows from the discussion in the third to next paragraph below.

A crucial determinant of β is σ_ε^2. If σ_ε^2 is low, β is close to one. Then voters' posterior beliefs about x_1^* come close to the true value and voters' prior beliefs have only little influence on their posterior beliefs. Therefore, μ_1^m has little weight in influencing the politician's choice of g_1 and the incentive to pander is only weak. To understand this, consider the hypothetical limit case where σ_ε^2 would approach zero. Then voters would perfectly observe x_1^* ex post. Clearly, the incumbent politician would then maximize the probability of getting reelected by setting $g_1 = \xi_1$, since $E\left[x_1^* \left| \xi_1\right.\right]$ from Assumption 2.

In the opposite case where σ_ε^2 is large, β is low. Thus, voters' beliefs are highly persistent and μ_1^m has a high weight in influencing g_1. As already mentioned in Section 3.1, a high σ_ε^2 may be understood as corresponding to ideological beliefs that voters may be motivated to keep intact (Bénabou, 2008). As a result, they blame ε for any observed outcome that deviates from their expectations.

In the limit case where $k = \infty$, (part (ii) of Proposition 1), policy making is perfectly populist and neither politician type makes use of her signal. In this case, the function $G^{-1}(g_1)$ does not exist since $\bar{\xi}_1$ cannot be inferred from g_1. This is the reason why off-equilibrium beliefs are required. An obvious supporting off-equilibrium belief is that a politician is incompetent whenever $g_1 \neq \mu_1^m$.[9]

[9] For $k = \infty$, there exists a continuum of pooling equilibria for suitably chosen off-equilibrium beliefs. We consider the equilibrium stated in Proposition 1 as the natural candidate since it provides the limiting case of equilibria under level-k beliefs.

The result that a higher k leads to more populism is best understood from a formal point of view. Assume that k is infinite. Then, it is in fact a logical impossibility that g_1 can depend on ξ_1. Suppose it would. Then voters would be aware of this. They also understand that $\xi_1 = x_1^*$ in the case of the competent politician. However, they do not observe x_1^* and voter i substitutes $\hat{\mu}_1^i$ for x_1^*. Thus, a politician who wants to appear competent to the median voter will not actually want to let his policy depend on ξ_1 but rather on $E\left[\hat{\mu}_1^m \left|\xi_1\right.\right] = (1-\beta)\,\mu_1^m + \beta\xi_1$. Here ξ_1 enters only with a weight β which lies between zero and one. But now g_1 would still depend on ξ_1, hence the same argument can be repeated and we would find that g_1 could in fact only depend on $(1-\beta^2)\,\mu_1^m + \beta^2\xi_1$. This argument can be iterated an infinite number of time. Since $0 < \beta < 1$, ξ_1 must necessarily vanish and g_1 cannot depend on ξ_1. If k is finite, then this argument can be repeated only a finite number of times that increases with k. With each iteration, g_1 depends less on ξ_1 and more on μ_1^m since $0 < \beta < 1$.

Now we briefly consider the second period. Our assumption is that the politician in office in the second period maximizes welfare since she does not have any incentives to manipulate voters' beliefs.[10] We summarize the findings for the second period in the following proposition.

Proposition 2 (Equilibrium Second Period) *(i) In the second office period, a competent politician always sets $g_2 = x_2^*$. An incompetent reelected incumbent sets $g_2 = x_2^* + \zeta_1$ and an incompetent challenger sets $g_2 = x_2^* + \zeta_2$. (ii) The probability that a competent incumbent gets reelected is at least as high as the probability that an incompetent incumbent gets reelected. (iii) The probability that the second period politician is competent is greater or equal to α.*

[10] The politician may still have goals other than maximizing welfare and may engage in rent extraction. In principle, we could allow for rent seeking behavior in a way similar to Persson and Tabellini (2000, Ch. 4.5). Suppose that there is an upper bound on the amount of rents that a politician can extract. Suppose further that a competent politician makes better use of the remaining government budget by better promoting welfare due to superior information. Then all substantive conclusions of our model would still hold. The only difference would be that welfare in the second period is reduced by an identical amount for both politician types due to the loss experienced from rent extraction.

4 The Costs and Benefits of Populist Policies

Using Proposition 1 and 2, it is straightforward to characterize welfare under indirect democracy. We will do so by using the concept of a loss function defined as $L_t = EV_t^{FB} - EV_t^{EQ}$ for office period t. L_t is defined as the difference between expected utility as achieved when g_2 is set to its ex ante welfare-maximizing level and expected utility as achieved in the equilibrium of the political game. The first-best utility value ex ante is obtained by setting $g_t = x_t^*$ which leads to $EV_t^{FB} = -\sigma_\varepsilon^2$. We obtain:

Proposition 3 (Welfare Indirect Democracy) *Under indirect democracy, welfare is characterized by*

$$L_1^{ID} = \left(1-\beta^{k+1}\right)^2 (x_1^* - \mu_1^m)^2 + \beta^{2(k+1)} (1-\alpha) \sigma_\zeta^2. \tag{8}$$

$$L_2^{ID} = [1 - \alpha (1 + \Delta_\alpha)] \sigma_\zeta^2, \tag{9}$$

where $\Delta_\alpha \geq 0$.

Consider the first period. The welfare loss from indirect democracy is equal to a weighted average of the median voter's bias and the variance of the incompetent politician's signal. The first term arises from pandering. The second term arises from the fact that no equilibrium for finite k entails full pandering. Politicians will always partially base their policy choice upon their signal ξ_1. This follows from the fact that voters observe $x_1^* + \varepsilon_1$ and update their beliefs about x_1^* and hence the welfare maximizing level of g_1 before they make their reelection decision. Anticipating this induces the incumbent politician to not fully ignore her signal. Since the signal of the incompetent politician is noisy, this increases the variance of g_1 which comes at a cost. Note that the weights $\left(1-\beta^{k+1}\right)^2$ and $\beta^{2(k+1)}$ do not add to one. We will come back to this in the next section.

It follows from Proposition 3 that there are both costs and benefits to populist policies. The costs relate to the fact that the competent politician partially ignores her signal. Since the signal reveals the truth this means wasting useful information. On the other hand, populism also leads the incompetent politician to partially ignore her signal. This may come at a benefit if $\sigma^2_{1\zeta}$ is relatively large and it may prevent policy making from being too erratic. The key insight from this is that populism can have beneficial consequences in a world where both voters and policy makers are imperfectly informed. This is a crucial insight for judging the benefits of democracy.

In the second period, populism does not arise since no politician has an incentive to manipulate voters' perception of their competence. As a result, only the noise term σ^2_ζ contributes to the welfare loss.

5 Direct Democracy and Non-Accountable Agents

We now turn to the second important question of this paper: which political institutions are optimal if voters are imperfectly informed about the world, but politicians and other agents may be so, too.

We first consider direct democracy. We follow Maskin and Tirole (2004) by modeling direct democracy as a political setup where $g_t = \mu^m_t$, i.e. the policy is set according to the median voter's prior beliefs. The idea is that in a direct democracy voters have the right to ask for referenda and that this would lead to a strong link between policy making and the beliefs of the median voter.[11] The following proposition follows directly from inserting g_t into (1) and taking expectations.

Proposition 4 (Welfare under Direct Democracy) *Under direct democracy,* $L^{DD}_t = (x^*_t - \mu^m_t)^2$.

The loss function is again defined as the deviation of expected utility from its first-best level.

Before we compare this to the case of indirect democracy, we introduce the third system that we consider here: delegation of policy making to independent agents. In the following, we will dub these agents experts, since this reflects more accurately what we have in mind.

[11] Switzerland provides an example of a country where voters have the right to ask for a referendum by submitting a petition signed by at least 100,000 registered voters.

In order to facilitate the comparison to direct democracy, our assumptions about experts parallel our assumptions about politicians. In particular, we also assume that there are two types of experts, competent and incompetent. Exactly as in the case of politicians, we assume that the competent politician receives a signal, denoted ξ_t^e, about x_t^*. For the competent expert we have $\xi_t^e = x_t^*$ while, for the incompetent expert, we have $\xi_t^e = x_t^* + \nu_t$. The random variable ν_t reflects a noise term with an expected value of zero and a variance σ_ν^2. The probability that an expert is competent is π. By definition, experts as non-accountable agents conduct policy for both periods and cannot be fired after the first period.

The combination "incompetent expert" may sound rather odd at first. What we have in mind is that if experts disagree, at most one expert opinion can be right. Thus, experts may be wrong even if they are highly trained simply as a result of the fact that human knowledge is highly imperfect, even at the level of top experts. Combating crime provides one salient example where experts disagree substantially (see Levitt, 1998, and Buscaglia, 2008), climate change provides another one (see McKibbin and Wilcoxen, 2002; Weitzman, 2007; Stern 2008).

In one important aspect, our assumptions about experts deviate from the assumptions made about politicians. We assume that experts are fully benevolent. Thus, they set $g_t = \xi_t^e$. We make this assumption since the case of benevolent experts is often considered as an ideal, if unfeasible, benchmark for government. Here we are interested in the question under which conditions this ideal benchmark would actually be desirable in a world of imperfect knowledge.

The welfare loss under the expert system is given in the following proposition. The proof is very similar to the one of Proposition 3 and is omitted.

Proposition 5 (Welfare Experts) *In the case of non-accountable experts,* $L_t^{EXP} = (1-\pi)\,\sigma_\nu^2$.

We are now in a position to compare the three political institutions. We start the discussion by noting that, concerning the first period, indirect democracy can be understood as a mix of direct democracy and non-accountable experts. Denote ξ_1^p the signal of a politician, and assume, for the sake of system comparison, that $\alpha = \pi$ and $\xi_1^e = \xi_1^p \equiv \xi$. Then we have $g_1^{DD} = \mu_1^m$,

$g_1^{EXP} = \xi$, where DD refers to direct democracy and EXP to experts. From Proposition 1 it follows that $g_1^{ID} = \left(1 - \beta^{k+1}\right) g_1^{DD} + \beta^{k+1} g_1^{EXP}$ in this case (where ID refers to indirect democracy).

This weighted-average nature of indirect democracy makes it attractive to risk averse voters in the sense that $L_1\left(g_1^{ID}\right) < \left(1 - \beta^{k+1}\right) L_1\left(g_1^{DD}\right) + \beta^{k+1} L_1\left(g_1^{EXP}\right)$. This follows from the fact that L_1 is strictly convex. The fact that the loss associated with g_1^{ID} is lower than a weighted average of the losses associated with either g_1^{DD} or g_1^{EXP} is also the reason why the weights associated with the two terms in L_1^{ID} in Proposition 3, namely $\left(1 - \beta^{k+1}\right)^2$ and $\beta^{2(k+1)}$, add to less than one for finite k. We summarize this finding in a corollary.

Corollary 1 (Comparative Advantage of Indirect Democracy) *Suppose that* $\alpha = \pi$ *and* $\xi_1^e = \xi_1^p$. *Then* $g_1^{ID} = \left(1 - \beta^{k+1}\right) g_1^{DD} + \beta^{k+1} g_1^{EXP}$ *and* $L_1\left(g_1^{ID}\right) < \left(1 - \beta^{k+1}\right) L_1\left(g_1^{DD}\right) + \beta^{k+1} L_1\left(g_1^{EXP}\right)$.

Comparing the loss functions in Proposition 3, 4, 5, the elements that crucially affect which institution is optimal are: The bias of the median voter's belief $|\mu_t^m - x_t^*|$; the variance of the incompetent politician's signal σ_ζ^2; and the corresponding variance of the expert's signal σ_ν^2. We refer a discussion of the meaning of these parameters and the practical implications of our analysis to the next section. Here, we simply aim to point out the following conclusions:

- Non-accountable experts are optimal if σ_ν^2 is small relative to $|\mu_t^m - x_t^*|$ and σ_ζ^2.
- Direct democracy is optimal if $|\mu_t^m - x_t^*|$ is small relative to σ_ζ^2 and σ_ν^2.
- In case that neither of these conditions applies, the weighted-average nature of indirect democracy may often make it optimal.

6 Discussion

We first provide arguments why the variance of the noise of experts' signal σ_ν^2 may not be small in many cases. This leads us to the conclusion that delegation to experts may not be optimal in many domains. We proceed by discussing the relative merits of direct and indirect democracy. Our conclusion is that the latter may be optimal in many instances.

The noise of experts' signal σ_ν^2 reflects the degree to which experts can draw on reliable knowledge that has been derived from a large amount of high-quality data. An example of a piece of knowledge associated with a very low σ_ν^2 would be "HIV causes AIDS."

The desirability of many policies depends on *behavioral reactions* to certain interventions, e.g. the reactions of criminal activities to more severe punishment, or the reactions of labor supply to a five-percent increase in the income tax. Unfortunately, there are few examples where social scientists can draw upon sharp empirical knowledge in order to pin down the consequences to a policy intervention with a high degree of precision. This is a consequence of two fundamental facts. First, it is rarely possible to conduct randomized experiments in real-world setups.[12] As a result, the availability of high-quality data is naturally limited. Second, many important determinants of behavior are very hard to measure and will never be well observed. These include preferences, beliefs, personality characteristics, management skills etc., in short what is often dubbed unobserved heterogeneity. Third, many policy interventions have a unique element in that they are carried out for the first time or for the first time under particular circumstances or in a particular country. There may thus be limited opportunity to predict behavioral reactions based on existing data. Beyond this, there is also a large degree of fundamental uncertainty in domains of *natural sciences* that are highly relevant for policy making, most notably in the case of climate change.

As a result of these fundamental facts, σ_ν^2 is probably not small in many important cases. In line with this observation, there are many examples where experts widely disagree (e.g. combatting crime or climate change; see references above). In these cases, it is conceivable that the median voter's prior belief lies somewhere in between the signals of various experts.

In light of the fact that scientific knowledge about the behavioral effects of policy interventions may often be highly incomplete, society may want to base policy decisions partly on

[12]For exceptions see Banerjee and Duflo (2008).

common sense and practical judgment. This may allow for reducing the extent of errors made from basing policy decisions purely on scientific knowledge which is necessarily highly incomplete. Unfortunately, experts may often not be expected to be particularly good at applying common sense judgment since they may often have been less exposed to real-world experiences than either politicians or voters themselves. This is related to the effect of group thinking (Sunstein, 2001), i.e. the observation that people's view become distorted as a result of interacting mainly with other people from a similar group. Experts are highly trained intellectuals who are likely to interact mainly among themselves. As a result, it is conceivable that their ability to make practical judgment about expected responses to a policy intervention is less reliable than either in the case of politicians or voters. This discussion relates to our model in that the potential "ivory-tower" nature of experts' judgments increases σ_ν^2 beyond the level associated with just imperfect empirical knowledge.

In contrast, σ_ν^2 is low and hence delegation to experts desirable in domains where choosing the right policy is simply a matter of applying technical knowledge. Here, we come to the same conclusions as Maskin and Tirole (2004), and Alesina and Tabellini (2007, 2008). As an example, we mention the calculation of an annuity conversion factor for a public pension system as a function of life expectancies. Central banking may be another important case.

Now we turn to the comparison of direct and indirect democracy. Our analysis implies that direct democracy is desirable if the bias of the median voter's prior $|\mu_t^m - x_t^*|$ is small. There are reasons why this term should be expected to be small in some circumstances and large in others. It is expected to be small when judging the effect of a policy does not so much require formal knowledge but rather a good "feeling" about what would be an appropriate policy action and what the likely behavioral reactions are. Second, in some cases, voters may actually have more information than politicians (and experts). An example for the former case may be a judgment about how much free choice of savings and investment strategies there should be in a pension system. Ordinary voters may be better able to judge how easy or difficult it may be for them to make such choices themselves than politicians (and experts). An example for the latter case may be combating crime in local neighborhoods. In both cases, $|\mu_t^m - x_t^*|$ may be small relative to σ_ζ^2 and σ_ν^2.

In other instances, judging the desirability of a policy may require more formal knowledge. An important example may be globalization, although even here ordinary workers may have some additional information about how they are affected by globalization in the short-run that is not readily available to politicians and experts. To the degree that formal knowledge is important for judging policies, politicians have greater incentives to acquire this knowledge than voters (Maskin and Tirole, 2004). The reason is that the latter may anticipate that they are not pivotal and thus not invest in the acquisition of this formal knowledge. This makes the case for delegating policy making to politicians. On the other hand, there is the danger that politicians may be ideological (as may be voters) or ignore informal information, which is reflected in the parameter σ_ζ^2. This is where the positive effects of "populism" come into play: politicians are induced not to rely exclusively on their signal.

Our main positive result in this paper is that policies under indirect democracy are determined as a weighted average of voters' priors and politicians' signals. This weighted average nature makes indirect democracy a very balanced and well-diversified institution which is desirable from a normative point of view. Our analysis may help to explain why it is so prevalent around the world.

7 Conclusion

In this paper, we have analyzed policy making under indirect democracy when voters have only imperfect knowledge about the world. We have investigated the welfare implications of the fact that politicians have an incentive to pander to voters' beliefs. Furthermore, we have addressed the question whether direct democracy or delegation of policy making to non-accountable agents may improve welfare compared to indirect democracy. Our analysis has been carried out in a setup where also politicians and experts in the role of non-accountable agents have only imperfect knowledge about the world.

Our main positive result is that policies under indirect democracy are a weighted average of voters' prior beliefs and politicians' signals. This implies our main normative result that, due to its balanced nature, indirect democracy should often be expected to be preferable to either direct democracy or delegation to non-accountable agents. An important exception is the case where policy making mainly requires application of technical knowledge.

There is a range of issues to be addressed in future research. First, it would interesting to generalize the model to an indefinite time horizon where each politician may serve in office for two or more consecutive periods. Second, the objective function of politicians may be enriched by other goals such as acquiring rents or carrying out prestigious projects. A difficult but interesting topic would be to consider the case where a policy affects welfare of different subgroups of the population in different ways. It would also be important to think about which forms of indirect democracy may be most desirable, either presidential or parliamentary. Finally, the setup of this paper may also apply to decision making in corporations. Both managers and owners may only have imperfect information about profit-maximizing actions but managers' careers may depend on the judgement of owners. Therefore, managers could be inclined to act in line with owners' beliefs and purposefully neglect their own (specialist's) signal on the profit-maximizing action.

Appendix A: Proofs

Proof of Lemma 1

Voters observe $x_1^* + \varepsilon_1$. From the point of view of voter i, $x_1^* + \varepsilon_1$ is a realization of the random variable $x_1^i + \varepsilon_1$. The voter aims to update his belief about x_1^i. The random variables x_1^i and $x_1^i + \varepsilon_1$ are jointly normally distributed with $E\left[x_1^i\right] = \mu_1^i$, $Var\left[x_1^i\right] = \sigma_{1x}^2$, $E\left[x_1^i + \varepsilon_1\right] = \mu_1^i$, $Var\left[x_1^i + \varepsilon_1\right] = \sigma_{1x}^2 + \sigma_\varepsilon^2$. Furthermore, $Cov\left[x_1^i, x_1^i + \varepsilon_1\right] = \sigma_{1x}^2$. Inserting this in the formulas for conditional expectations and variances for jointly-normal random variables (see e.g. Hogg and Craig, 1995, p. 148) yields the result.

Proof of Lemma 3

We first prove (2). For simplicity, we omit the time subscript as well as the superscript i when there is no danger of confusion. The idea of the proof is to derive the posterior probability $\hat{\alpha}$ for the case that the random variable ξ falls into the (small) interval $I_\delta\left(\bar{\xi}\right) := \left[\bar{\xi} - \delta,\ \bar{\xi} + \delta\right]$ and to consider the limit $\delta \to 0$.

Denote by C the event that a politician is competent and by IC the complementary event. Using the definition of conditional probabilities, it follows that

$$P\left(C\left|\xi \in I_\delta\right.\right) = \frac{\alpha P\left(\xi \in I_\delta\left|C\right.\right)}{\alpha P\left(\xi \in I_\delta\left|C\right.\right) + (1-\alpha) P\left(\xi \in I_\delta\left|IC\right.\right)}. \tag{10}$$

Note that the denominator is equal to $P\left(\xi \in I_\delta\right)$. In order to consider the limit of (10) for the case where $\delta \to 0$, it is useful to rewrite it as

$$P\left(C\left|\xi \in I_\delta\right.\right) = \left[1 + \frac{1-\alpha}{\alpha} \frac{\int_{\bar{\xi}-\delta}^{\bar{\xi}+\delta} f_{ic}\left(\xi\right) d\xi}{\int_{\bar{\xi}-\delta}^{\bar{\xi}+\delta} f_c\left(\xi\right) d\xi}\right]^{-1}. \tag{11}$$

Since the two normal density functions f_c and f_{ic} are well-behaved, it follows from standard arguments using the definition of the Riemann integral that $\lim_{\delta \to 0} \frac{\int_{\bar{\xi}-\delta}^{\bar{\xi}+\delta} f_{ic}(\xi) d\xi}{\int_{\bar{\xi}-\delta}^{\bar{\xi}+\delta} f_c(\xi) d\xi} = \frac{f_{ic}\left(\bar{\xi}\right)}{f_c\left(\bar{\xi}\right)}$. Substituting this into (11) and rearranging yields (2).

We now prove that $\hat{\alpha}$ is decreasing in $\left|\bar{\xi}-\hat{\mu}_1\right|$. We consider the case that $\bar{\xi}-\hat{\mu}_1 \geq 0$. Write $\hat{\alpha}=\left[1+\frac{1-\alpha}{\alpha}\frac{f_{ic}\left(\bar{\xi}\right)}{f_c\left(\bar{\xi}\right)}\right]^{-1}$. Since $\hat{\mu}_1$ is to be treated as a constant here, it is sufficient to show that $\frac{f_{ic}\left(\bar{\xi}\right)}{f_c\left(\bar{\xi}\right)}$ increases with $\bar{\xi}$. f_c is the normal density describing the distribution of $\xi_c \equiv \hat{x}_1$, while f_{ic} is the normal density associated with $\xi_{ic} \equiv \hat{x}_1+\hat{\zeta}_1$. Using the formula for the normal density, we have $f_c\left(\xi\right)=\frac{1}{\sqrt{2\pi Var(\xi_c)}}\exp\left[-\frac{(\xi-E\xi_c)^2}{2Var(\xi_c)}\right]$ and $f_{ic}\left(\xi\right)=\frac{1}{\sqrt{2\pi Var(\xi_{ic})}}\exp\left[-\frac{(\xi-E\xi_{ic})^2}{2Var(\xi_{ic})}\right]$. It then follows that

$$d\left[\frac{f_{ic}\left(\xi\right)}{f_c\left(\xi\right)}\right]/d\xi=\frac{\sqrt{Var\left(\xi_c\right)}}{\sqrt{Var\left(\xi_{ic}\right)}}\left[\frac{\xi-E\xi_c}{Var\left(\xi_c\right)}-\frac{\xi-E\xi_{ic}}{Var\left(\xi_{ic}\right)}\right]\exp\left[\frac{\left(\xi-E\xi_c\right)^2}{2Var\left(\xi_c\right)}-\frac{\left(\xi-E\xi_{ic}\right)^2}{2Var\left(\xi_{ic}\right)}\right].$$

Since we consider the case that $\xi=\bar{\xi}\geq E\xi_c=\hat{\mu}_1$, and since $\bar{\xi}=G^{-1}\left(g_1\right)$, Lemma 2 implies that $\bar{\xi}\geq E\xi_{ic}\geq E\xi_c$. Furthermore, $Var\left(\xi_{ic}\right)>Var\left(\xi_c\right)$. These inequalities imply $d\left[\frac{f_{ic}\left(\bar{\xi}\right)}{f_c\left(\bar{\xi}\right)}\right]/d\bar{\xi}>0$.

The case where $\bar{\xi}<\hat{\mu}_1$ is symmetric and analyzed by following the same steps.

Proof of Lemma 5

Proof of (i). If $\mu_1^i=\mu_1^m$ for at least half of voters then it follows immediately that $\bar{\xi}_1 \in I_{re}^m$ implies that $\bar{\xi}_1 \in I_{re}^i$ for at least half of voters. Conversely, $\bar{\xi}_1 \notin I_{re}^m$ implies that $\bar{\xi}_1 \notin I_{re}^i$ for at least half of voters.

Proof of (ii). It follows from Lemma 1 that condition $\mu_1^{\max}-\mu_1^{\min}\leq 2\delta^{crit}/\left(1-\beta\right)$ is equivalent to $\max I_{re}^{\min}\geq \min I_{re}^{\max}$, where the superscripts min and max refer to the i with the lowest and highest μ_1^i, respectively. If this holds then $\bar{\xi}_1\in I_{re}^m$ implies that either $\bar{\xi}_1\in I_{re}^{\max}$ or $\bar{\xi}_1\in I_{re}^{\min}$. Furthermore, whenever $g\in I_{re}^m$ and $g\in I_{re}^{\min}$, then also $g\in I_{re}^i$ for all i with $\mu_1^{min}\leq\mu_1^i\leq\mu_1^m$. Similarly, whenever $g\in I_{re}^m$ and $g\in I_{re}^{\max}$, then also $g\in I_{re}^i$ for all i with $\mu_1^m\leq\mu_1^i\leq\mu_1^{\max}$. Hence, in either case, if $\bar{\xi}_1\in I_{re}^m$ then $\bar{\xi}_1$ will belong to the reelection interval of at least half of voters.

Consider now the case that $\bar{\xi}_1\notin I_{re}^m$. It follows that either $\bar{\xi}_1\notin I_{re}^i$ for all i with $\mu_1^{min}\leq$

$\mu_1^i \le \mu_1^m$ or for all i with $\mu_1^m \le \mu_1^i \le \mu_1^{\max}$. In either case, $\bar{\xi}_1 \notin I_{re}^i$ holds for at least 50 percent of the voters. Overall, this establishes that m is pivotal.

Proof of Proposition 1

Proof of part (i). From Definition 1, the median voter's belief is that a politician sets $g_1 = G(\xi_1) = \mu_1^m + \beta^k(\xi_1 - \mu_1^m)$. By Lemma 6, (6) and (7) are a politician's best response to these beliefs. The equilibrium is unique since beliefs are well-defined for any real-valued ξ_1 and g_1. This follows from the fact that ξ_1 is normally distributed for both politician types, which is known by voters. Hence, the equilibrium does not hinge on off-equilibrium beliefs.

Proof of part (ii). For infinite k Lemma 6 does not apply. The pooling equilibrium where both types set $g_1 = \mu_1^m$ is obtained as a limit case of (6) and (7) for $k \to \infty$. This equilibrium is supported by appropriate off-equilibrium beliefs, e.g. the belief of all voters that a politician setting $g_1 \neq \mu_1^m$ is incompetent with probability one. Given this off-equilibrium belief, setting $g_1 = \mu_1^m$ maximizes the probability of getting reelected. This follows from Lemma 4. For $g_1 \neq \mu_1^m$ we have that $\hat{\alpha}^i = 0$ for all i. Furthermore, $E\hat{\zeta}_1^i = 0$ and $\hat{\sigma}_{1\zeta}^2 = \sigma_\zeta^2$ since it is not possible to update beliefs about ζ_1. Thus, the left-hand side of (5) is larger than one. In contrast, it is equal to one for $g_1 = \mu_1^m$ since voters are indifferent between reelecting our ousting the incumbent.

Proof of Proposition 2

Proof of part (i). This follows directly from the assumption that the politician in office in the second period maximizes welfare.

Proof of part (ii). We have to show that the probability that (5) holds for the median voter is lower for an incompetent incumbent than for a competent incumbent. We first prove this for finite k. By Lemma 3, $\hat{\alpha}^m$ is a strictly decreasing function of $\left|\bar{\xi}_1 - \hat{\mu}_1^m\right|$. By Definition 1, the median voter's belief is that $g_1 = G(\xi_1) = \mu_1^m + \beta^k(\xi_1 - \mu_1^m)$. Hence, $\bar{\xi}_1 = \frac{g_1}{\beta^k} - \frac{1-\beta^k}{\beta^k}\mu_1^m$. By Proposition 1, $g_1 = \left(1-\beta^{k+1}\right)\mu_1^m + \beta^{k+1}\xi_1$. In the case of the competent politician, $\xi_1 = x_1^*$.

Inserting this into the expression for g_1, inserting then into the expression for $\bar{\xi}_1$ and using Lemma 1 yields that $\bar{\xi}_1 - \hat{\mu}_1^m = -\beta\varepsilon_1 \equiv \phi_c$. Similarly, it follows for the incompetent politician that $\bar{\xi}_1 - \hat{\mu}_1^m = \beta(\zeta_1 - \varepsilon_1) \equiv \phi_{ic}$. The variance of ϕ_{ic} is equal to $\beta^2 \left(\sigma_{1\zeta}^2 + \sigma_\varepsilon^2\right)$, whereas the variance of ϕ_c is equal to $\beta^2\sigma_{1\zeta}^2$ and thus strictly smaller than the variance of ϕ_{ic}. It follows that the probability that $\left|\bar{\xi}_1 - \hat{\mu}_1^m\right| \geq A$, for any $A \in \mathbf{R}_{++}$, is strictly greater for the incompetent than for the competent politician. Hence, the probability that $\hat{\alpha}^m \leq B$, for any $B \in (0,1)$, is strictly greater for the incompetent than for the competent politician.

From Lemma 2, $E\hat{\zeta}_1^m = \gamma\left(\bar{\xi}_1 - \hat{\mu}_1^m\right)$. (Note that $G^{-1}(g_1) = \bar{\xi}_1$.) Thus, the above arguments also imply that the probability that $\left(E\hat{\zeta}_1^m\right)^2 \geq C$, for any $C \in \mathbf{R}_{++}$, is strictly greater for the incompetent incumbent than for the competent incumbent. This establishes that the probability that (5) holds for the median voter is strictly smaller in case of an incompetent incumbent than in case of a competent incumbent for finite k.

If k is infinite, $\bar{\xi}_1$ cannot be inferred and no information about the incumbent's type is observed since we have a pooling equilibrium. Hence, the probability of getting reelected must be equal for both types.

Proof of part (iii). There are three events in which the politician in the second period is competent: (1) A competent incumbent gets reelected; (2) a competent incumbent gets ousted and replaced by a competent politician; (3) an incompetent incumbent gets ousted and replaced by a competent politician. Denote the probability that a competent politician gets reelected by ρ_c and the probability that an incompetent politician gets reelected as ρ_{ic}. Denote the event that the second period politician is competent by C_2. We have then

$$\begin{aligned} Pr\left[C_2\right] &= \alpha\rho_c + \alpha^2(1-\rho_c) + \alpha(1-\alpha)(1-\rho_{ic}) \\ &= \alpha\left[1 + (1-\alpha)(\rho_c - \rho_{ic})\right] \geq \alpha. \end{aligned} \tag{12}$$

The last inequality follows from using the result of part (ii).

Proof of Proposition 3

Denote by $g_{t,c}$ the level of g set in period t by the competent politician and let $g_{t,ic}$ refer to the incompetent politician. Let ρ_t denote the probability that a politician is competent in period t. Then

$$EV_t = -(\rho_t E\left(g_{t,c} - x_t^* - \varepsilon_t\right) + (1 - \rho_t)\, E\left(g_{t,ic} - x_t^* - \varepsilon_t\right)). \tag{13}$$

We have $\rho_1 = \alpha$. From Proposition 2, $\rho_2 = \alpha + \Delta_\alpha$ for some $\Delta_\alpha \geq 0$. Using this and inserting the expressions for $g_{t,c}$, $g_{t,ic}$ given in Proposition 1 and 2 into (13) yields

$$\begin{aligned} EV_1 &= -\left(1 - \beta^{k+1}\right)^2 \left(\mu_1^m - x_1^*\right)^2 - (1-\alpha)\, \beta^{2(k+1)} \sigma_\zeta^2 - \sigma_\varepsilon^2, \\ EV_2 &= -\left(1 - (\alpha + \Delta_\alpha)\right) \sigma_{1\zeta}^2 - \sigma_\varepsilon^2. \end{aligned}$$

EV_1 is maximized for $g_t = x_t^*$ which yields $EV_t^{FB} = -\sigma_\varepsilon^2$. Inserting this and the above expressions into the definition of L_t yields the result.

Appendix B: Indirect Democracy in the Case of Non-Degenerate Beliefs of Politicians

In the baseline version of our model of indirect democracy we assume that politicians receive a signal which directly characterizes their expected value of $E\left[x_1^*\right]$ (see Assumption 2). This introduces an asymmetry between voters and politicians since the former hold prior beliefs that x_1^* is distributed according to a random variable x_1^i.

Instead of assuming that a politician's signal directly determines $E\left[x_1^*\right]$, we could assume that the signal determines a politician's prior belief about x_1^*, as in the case of voters. Specifically, we may assume that a politician's beliefs about x_1^* are that it is distributed according to $x_1^p \sim N\left(\xi_1^p, \tilde{\sigma}_{1x}\right)$, where the superscript p indexes a politician. This would affect the analysis only insofar as a politician would be able to update her beliefs about x_1^* after observing $x_1^* + \varepsilon_1$, as do voters.

The consequences for our analysis are minor. Consider first the case where x_2^* is unrelated to x_1^* in the sense that $E\left[x_2^*\left|x_1^*\right.\right] = E\left[x_2^*\right]$. Here, the logic is that x_t^* is drawn by nature in each period and the expected values are with respect to the distribution that governs nature's draws. In this case, the possibility of updating her beliefs about x_1^* would not affect any of the politician's decisions. The reason is that g_1 has to be set before $x_1^* + \varepsilon_1$ is observed. Second, any information about x_1^* is not useful for predicting x_2^*.

Suppose now that, generically, $E\left[x_2^*\left|x_1^*\right.\right] \neq E\left[x_2^*\right]$, such that information about x_1^* is useful for predicting x_2^*. In this case, an incumbent politician reelected for the second period would take updated beliefs about x_1^* into account when setting g_2. Voters would anticipate this and this would affect their anticipated distribution of g_2 when set by the incumbent politician. However, the logic of the reelection decision is unaffected. It is still true that, by reelecting an incumbent, the voter has additional information compared to the case of a challenger. This reduces the uncertainty of g_2 and thus comes at a benefit, everything else equal. Thus, the logic behind (5) is unchanged, although the specific expression would be somewhat more involved.

Importantly, in any case, g_1 would still be determined such as to minimize the expected distance between $G^{-1}(g_1)$ and $\hat{\mu}_1^i$, given the politician's *prior* beliefs about x_1^*. It follows that g_1 is again determined as in Proposition 1. As a result, also the welfare loss in the first period L_1 in Proposition 3 is unaffected. The welfare loss in the second period L_t is affected in the sense that σ_ζ^2 would have to be replaced by an expression taking into account that an incumbent politician has the possibility to update her beliefs. This reduces the welfare impact of the noise associated with her signal.

Appendix C: Indirect Democracy for biased ζ_1

As mentioned in the main text, politicians may be understood as a representative sample of the general population if the political selection process is not biased in favor of the elite or any other particular group. If politicians are representative for the general population, then we would expect that the incompetent politician's signal is related to the distribution of voter's beliefs. (For the competent politician, this does not apply, since she observes the truth.) A simple way of capturing this is assuming that $E\zeta_t$ is related to $\mu_t^m - x_t^*$.

In the following, we consider the limit case where, from an objective point of view, $E\zeta_t = \mu_t^m - x_t^*$. By objective we mean from the point of view of the economic theorist analyzing the problem. In contrast, we need to assume that a politician believes that $E\zeta_t = 0$ for herself. Otherwise, she could make use of the information about $E\zeta_t$ to unbias her belief about x_t^*. Second, we also assume that voters believe that $E\zeta_t = 0$. More precisely, we assume here that a majority of voters hold beliefs that are identical to the beliefs of the median voter. In this case, it is appropriate to assume that a majority of voters believe that $E\zeta_t = 0$. Otherwise, their beliefs about $E\zeta_t$ would be inconsistent with their own beliefs about x_t^*.

In this case, all positive results in Section 3 and, in particular, Subsection 3.4 continue to hold. However, the welfare expressions in Proposition 3 are modified. In particular, we obtain

$$L_1 = \left[\alpha\left(1-\beta^{k+1}\right)^2 + 1 - \alpha\right](x_1^* - \mu_1^m)^2 + \beta^{2(k+1)}(1-\alpha)\,\sigma_\zeta^2. \tag{14}$$

$$L_2 = \left[1 - \alpha\left(1+\Delta_\alpha\right)\right]\left[(x_2^* - \mu_2^m)^2 + \sigma_\zeta^2\right], \tag{15}$$

As to be expected, the terms $(x_t^* - \mu_t^m)^2$ have a stronger influence on the welfare loss than in the baseline case. In particular, in the case of Proposition 3, the coefficient for $(x_1^* - \mu_1^m)^2$ is smaller than in the case of (14). The expression $(x_2^* - \mu_2^m)^2$ does not appear at all in Proposition 3. The overall conclusion is that if the incompetent politician's signal is biased towards the beliefs that are prevalent among voters, indirect democracy becomes more similar to direct democracy.

References

Alesina, Alberto, and Tabellini, Guido (2007), "Bureaucrats or Politicians? Part I: A Single Policy Task," *American Economic Review*, 97(1), 169-179.

Alesina, Alberto, and Tabellini, Guido (2008), "Bureaucrats or Politicians? Part II: Multiple Policy Tasks," *Journal of Public Economics*, 92, 426-447.

Banerjee, Abhijit V., and Duflo, Esther (2008), "The Experimental Approach to Development Economics," NBER Working Paper No. 14467.

Bénabou, Roland (2008), "Ideology," *Journal of the European Economic Association,* 6(2-3), pp. 321-352.

Buscaglia, Edgardo (2008), "The Paradox of Expected Punishment: Legal and Economic Factors Determining Success and Failure in the Fight against Organized Crime," *Review of Law & Economics*, 290-317.

Canes-Wrone, Brandice, Herron, Michael C., and Shotts, Kenneth W. (2001), "Leadership and Pandering, A Theory of Executive Policymaking," *American Journal of Political Science,* 45(3), 532-550.

Camerer, Colin F., Ho, Teck-Hua, and Chong, Juin-Kuan (2004), "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, 119(3), 861-898.

Caplan, Bryan (2001), "What Makes People Think like Economists? Evidence on Economic Cognition from the 'Survey of Americans and Economists on the Economy'," *Journal of Law and Economics*, 44, 395-426.

Hogg, Robert V., and Craig, Allen T. (1995), *Introduction to Mathematical Statistics,* Prentice Hall: New Jersey.

Levitt, Steven D. (1998), " Juvenile Crime and Punishment," *Journal of Political Economy*, 106(6), 1156-1185.

Maskin, Eric, and Tirole, Jean (2004), "The Politician and the Judge: Accountability in Government," *American Economic Review*, 94(4), 1034-1054.

McKibbin, Warwick J., and Peter J. Wilcoxen (2002), "The Role of Economics in Climate

Change Policy," *Journal of Economic Perspectives,* 16(2), 107-129.

Persson, Torsten, and Guido Tabellini (2000), *Political Economics: Explaining Economic Policy,* MIT Press.

Schultz, Christian (2008), "Information, Polarization and Term Length in Democracy," *Journal of Public Economics,* 92, 1078-1091.

Stern, Nicholas (2008), "The Economics of Climate Change," *American Economic Review: Papers & Proceedings*, 98(2), 1-37.

Sunstein, Cass (2001), *Designing Democracy: What Constitutions Do,* Oxford University Press, Oxford.

Weitzman, Martin L. (2007), "A Review of the Stern Review on the Economics of Climate Change," *Journal of Economic Literature,* 45(3), 703-724.