# Essays in Political Economy

Ole-Andreas Elvik Næss

DEPARTMENT OF ECONOMICS

NHH Norwegian School of Economics

A thesis submitted for the degree of

*Philosophiae Doctor (PhD)*

BERGEN 2020

# Acknowledgements

I would like to express my deepest appreciation to Eirik Gaard Kristiansen. As my supervisor, he has taught me extremely much about economics and conducting economic research. Eirik has spent countless hours reading different drafts of my papers, and he has always been available for meetings. Eirik has provided me with both constructive and motivating feedback, and he has treated me with the highest level of respect.

I am also extremely grateful to my second supervisor, Tore Ellingsen, for invaluable support on all chapters in my thesis. Our conversations, either in Bergen or over the telephone, have been inspiring and very useful.

I would also like to thank NHH for giving me the opportunity to pursue a PhD. In particular, I would like to thank Fred Schroyen, for making my TA work enjoyable, as well as Hans Jarle Kind and Justin Valasek and the other faculty members and staff at the Department of Economics.

I also appreciate all the new friends I have met at NHH. I would like to thank Andreas, Erling and Ingar for our great "support group", and I would like to extend a particular thanks to Ingar for helping me with a large number of topics, ranging from how-to-make-a-presentation to how-to-apply-for-exchange. I would also like to thank Felix, Ingvild, Sandra and the other PhD students at NHH for many good memories.

I am also grateful to Bergen kommune for cooperating with me on the electoral turnout project.

I would like to extend my deepest gratitude to my parents, Helene and Robert, for invaluable support. This work would not have been possible without them. I also want to thank my sister, Julie, and brother-in-law, Ulrik, for carefully reading and proofreading all chapters in my thesis.

I also wish to thank my grandparents. Agnar has taught me about science and history since I was a small child and inspired me to pursue an academic career, while Ruth always cared for me. Bjørg and Jan also provided a lot of support. Although she is not an economist, there are few people being as economical and following the rational consumer models as closely as Bjørg. I also want to thank the rest of my family, family-in-law and friends.

Last but not least, I would like to thank my wife Marie. She is not only the love of my life, but also my best friend and my everyday informal discussant. All the hours we have spent discussing economics have taught me a lot about the usefulness and limitations of economic research.

# Contents

# INTRODUCTION

We are living in a period of time where it is both interesting and rewarding to conduct social science research. Increasing inequality, the lack of response to climate change, political polarization, fake news and threats from non-democratic forces are just some of today's important issues belonging to the realm of the social sciences. In my thesis, I analyze and try to provide some insights into economical and political issues. I use a combination of economic theory and modern econometric techniques to answer three distinct questions in applied microeconomics, with a particular focus on political economy.

Two major trends observed in Western democracies over the past decades are increasing economic inequality and political polarization. In the first chapter of my thesis, I provide a link between these two trends. Another political trend is that many countries observe low and decreasing levels of electoral turnout. In the second chapter, I analyze whether a simple nudge in the form of a text message can increase the electoral turnout. In the third chapter of my thesis, I analyze under-reporting of crime. I provide a possible explanation to this important social issue and analyze the effect of welfare-increasing mechanisms.

The chapters in this thesis analyze different economic and political issues, but they share important underlying characteristics. When analyzing individual political and social choices, it is necessary to understand that the outcome of a choice potentially depends on the choices made by others. This reasoning is at the core of all three chapters in my thesis. I analyze how political parties and politicians make choices in settings where the electoral outcome also depends on the choices made by other parties. Similarly, the decision of whether or not to vote may depend on the voting of others. As in voting settings, the decision of whether or not to report a crime depends on the choices made by others, because it may be more likely to convict a criminal when other victims also choose to report.

I will now provide a short summary of the three chapters in my thesis.

**Chapter 1: A Theory of Right-Wing Populism** This chapter analyzes the relation between increasing inequality and political polarization. Over the past decades there has been a growth in economic inequality in Western countries (Piketty, 2015; Saez and Zucman, 2019). The standard voting models (Meltzer and Richard, 1981) predict that more inequality leads to more redistribution. The empirical support for this relation is not strong, but in many countries increasing inequality is correlated with a growth in political polarization (Bonica et al., 2013). This polarization is not only characterized by parties diverging in their positions on redistribution, but also implies polarization in other policy dimensions. There has been a growth of right-wing populism, and an observed association between right-wing economic policies and far-right non-economic policies (Akkerman, 2012).

This chapter proposes a possible explanation to these observations. I construct an electoral model where there is a left-wing party, which wants a high tax rate, and a right-wing party, that wants a low tax rate. The two parties propose linear tax rates and a policy on some second dimension.

When there is a low level of inequality, my model predicts policy convergence in both dimensions. When there is a high level of inequality, more voters want a high tax rate. This is bad news for the right-wing party, because it becomes difficult to win the election while proposing a low tax rate. However, this party can bundle a low tax rate with the minority position on the second dimension. This bundle creates a coalition between the voters that want a low tax rate and the voters with minority preferences on the second dimension. Polarization in economic preferences may hence have spillover effect to polarization in other policy dimensions.

This argument rationalizes why more inequality is associated with political polarization in multiple dimensions, and it may cast some light on the observed right-wing populism.

**Chapter 2: Population-Level Treatment Effects: Theory and Evidence from a Large-Scale Voting Experiment** In the second chapter of my thesis, I first analyze methods for drawing causal inference when a treatment is provided to an entire population. I argue that randomizing the timing of the treatment and providing the treatment to different parts of the population at marginally different points in time may allow for causal analysis.

I use this method to conduct a large-scale field experiment. Together with the local government in Bergen, Norway, I analyze the effect of a text message that encouraged people to vote in the 2017 Parliamentary election. The text message was sent at a random point in time some days before the election, and half of the districts received the text message a few hours before the rest of the city.

I have access to voting data that is measured precisely in time, and I find that voting immediately increased in Bergen after the text message. I use a regression discontinuity approach to show that the increase in voting in Bergen after the text message was unusually large. Comparing the first and second receivers of the message reveals the same pattern. After the first half of the city received the message, voting increased in this part relative to the other half of the city. I estimate a size of the treatment effect between 1.4 and 2 percentage points.

I also analyze the underlying mechanisms. A post-election survey asked voters about their opinions about the text message, and this survey indicates that the voters appreciated the informational content of the message. This means that the underlying mechanism can be interpreted as a noticeable reminder of the election (Dale and Strauss, 2009). Gerber and Rogers (2009) argue that such encouragements are more effective when they highlight descriptive social norms, but I do not find any evidence indicating that a social text message leads to a larger treatment effect.

**Chapter 3: Under-Reporting of Crime** The third chapter analyzes the incentives for crime reporting. Under-reporting of crime is widely regarded as a large problem (Luce et al., 2010). In this chapter, I provide an explanation for the observed under-reporting as well as an analysis of welfare-increasing mechanisms.

I construct a model where victims choose whether to report the crime. A report leads to conviction if there is sufficient evidence. For example, two reports may be necessary for conviction. A victim chooses to report if the expected benefits from reporting (which is the probability of conviction multiplied with the benefits from conviction) is larger than the costs of taking the case to court.

I show that there will be under-reporting of crime in equilibrium. The reason is that the victims fail to internalize the positive externalities associated with a crime report. Suppose there is a victim that marginally prefers not to report a crime. If this victim chooses to report, the report will also benefit other victims, because it becomes more likely that the other victims are able to convict their offenders.

I then proceed to analyze the effect of different measures to increase the incentives for crime reporting. In my model, there is a rational jury and some agents making false crime reports. This means that the jury, after observing a given number of reports, computes the probability that the defendant is guilty conditional on the number of observed reports. When analyzing the effect of different mechanisms, I must also take into account how such mechanisms affect the agents making false reports. When there are few strategic agents making false reports, I show that mechanisms that make the reporting process less costly and less uncertain will increase social welfare. In this setting, I show that the optimal mechanism is given by an Information Escrow (Ayres and Unkovic, 2012), where a report is delivered to a principal and is only taken to court if there is sufficient evidence for conviction. However, when there are more agents potentially willing to deliver false reports, I show that making the reporting process costly and uncertain may be necessary to avoid false reporting. The combination of these two features will have a stronger discouraging effect on false victims, and this argument can be used to rationalize several apparently sub-optimal features of reporting systems.

# Chapter 1

# A Theory of Right-Wing Populism

Ole-Andreas Elvik Næss [1]

**Abstract**

This paper presents an explanation to observed trends in redistribution, polarization and right-wing populism. I extend a one-dimensional voting model by taking into account that voters, in addition to having economic preferences, also care about cultural issues. Political parties diverge in the cultural dimension to be able to implement different tax rates, and this effect is particularly strong for the right-wing party. A higher level of pre-tax income inequality makes it more difficult to win the election by proposing a low tax rate, but the right-wing party may be able to win the election by bundling a low tax rate with a focus on cultural issues (*JEL*: D72, H20).

## 1.1 Introduction

This paper proposes a unified explanation for three puzzles concerning political developments in Western democracies over the last couple of decades. The first

[1] I would like to thank Lassi Ahlvik, Ingar Haaland, Bård Harstad, Hans Hvide, Kjetil Storesletten, Justin Valasek and seminar participants and discussants in Bergen, Munich and Rotterdam for helpful comments and discussions. I would like to extend a special thanks to Tore Ellingsen and Eirik Gaard Kristiansen for tremendous guidance and support. The usual disclaimer applies.

puzzle is that increased pre-tax income inequality has not been associated with higher tax rates. The second puzzle is that income inequality has been associated with increased polarization among legislators and parties, regarding cultural as well as economic policies. The third puzzle is that populist cultural policies are typically bundled with low-tax policies rather than high-tax policies.

The main contribution of this paper is to show that these trends can be explained by taking into account that voters have both economic and cultural preferences. I let strategic political parties propose policies in two dimensions, which will lead to interaction effects between changes in preferences in one dimension and the electoral outcome in both dimensions. An increase in economic inequality may for example imply that more voters want a high tax rate, which is bad news for a party that wants a low tax rate. However, by bundling a low tax rate with a focus on cultural issues, this party may attract both the voters that want low tax rates and the voters with strong cultural preferences.

The three empirical puzzles above are supported by convincing evidence. There has been a large increase in inequality over the last decades (Piketty, 2015; Piketty and Saez, 2003; Saez and Zucman, 2019). In democratic systems, more inequality ought to entail more redistribution if voters care exclusively about their own consumption. Yet, in many countries, higher inequality has been accompanied by constant or decreasing redistribution (Bonica et al., 2013; De Mello and Tiongson, 2006). The strong correlation between economic inequality and political polarization in the US has been documented by McCarty et al. (2016) and Duca and Saving (2016).

The growth of populist policies is associated with right-wing economic policies, both in the US and Europe.[2] Akkerman (2012) finds that far-right parties entered 27 governments between 1996 and 2010, and in *all* of these cases the far-right parties

[2]But it is clear that right-wing economic policies are not a *defining* characteristic of populist parties (Ivarsflaten, 2008; Mudde, 2010).

formed a coalition with economically right-wing parties. This trend has continued in recent years exemplified by governments of Austria, Finland and Norway, where far-right parties have entered into coalition governments with conservative parties.[3]

This paper constructs a two-dimensional voting model to explain these trends. Meltzer and Richard (1981) use the median voter framework developed by Black (1948) and Downs (1957) to analyze the incentives for redistribution, and they show that more economic inequality implies that both parties propose higher tax rates.[4] I add a second, binary dimension to this framework.[5] The second dimension is throughout this paper interpreted as some cultural dimension, and cultural populism may serve as an example of such a policy dimension.[6] A majority of voters prefer a certain cultural policy, but there is some uncertainty related to the exact size of this majority. A second departure from Meltzer and Richard (1981) is that I let the political parties care about the implemented tax rate in addition to winning the election.[7] The parties in this model propose a tax rate and a binary cultural policy. The parties want to maximize the probability of winning the election, and the left party wants to implement a high tax rate while the right party wants to implement a low tax rate.

If both parties propose the same cultural policy, then the parties will also converge to the tax rate preferred by the median voter in equilibrium. Both parties would prefer

[3]This pattern is also visible in the politics of the EU. A 2019 report from Corporate Europe Observatory for example shows that all far-right parties in the European Parliament voted against a minimum corporate tax rate of 25%. This report is available from https://corporateeurope.org/en/2019/05/europes-two-faced-authoritarian-right-anti-elite-parties-serving-big-business-interests.

[4]See also Romer (1975) and Roberts (1977).

[5]Poole and Rosenthal (1991) find that US roll-call voting can be explained primarily by the traditional left-wing axis and a second dimension. Enke (2018) finds that moral values influence electoral choices and that politicians may strategically supply such moral values.

[6]However, the concept of populism is generally difficult to define (Mudde and Kaltwasser, 2017).

[7]Wittman (1977) and Calvert (1985) model politicians to be policy-motivated. Empirical results about the importance of policy relative to electoral success are unclear. Fredriksson et al. (2011) find that politicians care mainly about winning, while Martin and Stevenson (2001) find that policy preferences may help to explain government formations.

a solution where the parties make small deviations towards their preferred tax rates without changing the winning probabilities. However, this cannot be an equilibrium in a one-dimensional game, as a small deviation towards the median tax rate will lead to a large increase in the winning probability. Proposing different cultural policies may enable the parties to propose different tax rates in equilibrium, because electoral victory is possible by bundling this preferred tax rate with a populist cultural policy. In this model, cultural policy divergence is chosen to enable different tax rates in equilibrium. The right party has more to gain from proposing the cultural policy preferred by a minority of voters for two reasons. Income (and wealth) is generally not symmetrically distributed around the median value of the distribution.[8] If the median voter prefers a strictly positive tax rate, it follows that a majority of voters gain from redistribution, which means that the losses for the rich voters are larger than the gains for the other voters. This means that rich voters may be more willing to compromise on cultural issues to get their preferred economic policy. This asymmetry may open up for a coalition between rich voters and voters preferring a populist cultural policy. When there is a high level of economic inequality, the median voter prefers a high tax rate. The right party is less satisfied with this outcome than the left party, which means that the right party has more to gain from cultural polarization.

When the right party chooses a minority cultural policy and a low tax rate, the left party will propose a tax rate that is strictly higher than the median tax rate, which will lead to a further increase in polarization. The reason is that the left party is also pushed towards the median voter's tax preferences in a one-dimensional game, but will choose a higher tax rate if differences in cultural policies make a higher tax rate possible.

[8] In my model, this asymmetry in income occurs even productivity is uniformly distributed, because more productive workers choose a higher level of effort.

The expected tax rate in the polarized equilibrium is lower than the median voter's preferred tax rate when there is a high level of inequality. The intuitive reason is that the right party only wants to propose a minority cultural policy if this leads to a relatively large probability of winning the election while proposing a low tax rate.

This model can explain the association between increasing inequality, redistribution, political polarization and right-wing populism. Interestingly, the model can also cast some light on other recently observed trends. Cultural polarization between political *parties* has been argued to be increasing more than the cultural polarization between *voters*.[9] This model rationalizes that cultural divergence may occur even in cases where cultural preferences are stable. A common explanation to the relationship between inequality and populism is that voters are suffering from economic insecurity and vote for populist policies to protest against the political elites, and this explanation has received substantial empirical support.[10] This finding may explain the origin of anti-elite preferences, but it does not explain the link with right-wing economic policies. A poor voter with anti-elite preferences would prefer a populist party that proposes anti-elite policies *and* a high level of redistribution. This paper argues that the right-wing party has more to gain from catering to these preferences.

I show that the results are robust to changes in the set-up of the model. I find similar effects when the parties are purely policy-motivated and care about the tax rate implemented by the other party. I also show that the results hold for certain levels of diminishing marginal utility of income. The right-wing party is more willing to propose a diverging cultural policy when there is a high level of inequality for *two* reasons; more inequality increases the median tax rate and the richest voters are more

[9] See e.g., Fiorina and Abrams (2008), Bertrand and Kamenica (2018) and Desmet and Wacziarg (2018). Comparing data from the European Social Survey (ESS) in 2002 and 2014 suggests that preferences about important cultural dimensions are relatively stable. Europeans are for example on average becoming slightly more positive about immigration (Heath and Richards, 2016). Sub-national studies show the same pattern (Dennison and Talò, 2017).

[10] See e.g. Autor et al. (2016), Algan et al. (2017), Guiso et al. (2017) and Dal Bó et al. (2019).

willing to accept a non-preferred cultural policy. Diminishing marginal utility of income may dampen the second effect, but will generally have an ambiguous effect on the median tax rate. In the main model, the voters' tax preferences arise from a model of endogenous labor supply, similar to the model by Meltzer and Richard (1981). I also construct a simpler model with exogenous tax preferences, and I show that the results also hold within this framework.

In the main model, there are only two parties, but I also extend the model to include a third party that has preferences over the cultural policy dimension. I show that there will be a coalition between the third party and the right party unless the third party is too extreme, in which case the outcome is given by minority rule or a coalition between the right and left parties. Remarkably, this result shows that the underlying logic also holds when the parties are motivated by non-economic preferences. A cultural policy that is preferred by a minority of voters will not prevail if the cultural policy is the only policy dimension. However, parties preferring a minority cultural policy may use the divergence in economic preferences to create a coalition with rich voters that may be able to win the election while proposing a minority cultural policy. This argument implies that more economic inequality creates an opportunity to implement a cultural policy that is preferred by a minority of the voters by exploiting the divergence in the voters' tax preferences. This reasoning can cast some light over recently observed electoral coalitions in multi-party systems.

**Related literature** A large literature presents both theoretical and empirical arguments to explain the relationship between inequality and redistribution.[11] Shayo (2009) presents a model of redistribution with endogenous social identities, while

[11] See e.g., Benabou (2000), Benabou and Ok (2001), Rodriguez (2004), Fernández and Levy (2008), Karabarbounis (2011), Kuziemko et al. (2015), Barth et al. (2015) and Karadja et al. (2017).

strategic extremism is analyzed by Glaeser et al. (2005).[12] The multi-dimensional redistribution literature is reviewed by Iversen and Goplerud (2018).[13] The argument presented in this paper is related to the Machiavellian concept of "divide-and-conquer" analyzed by Marx, who suggested that racism is being used by capitalists to divide the working class.[14] Edsall and Edsall (1992) and López (2015) find that similar arguments have been used to divide the working class in the US. Frank (2004) argues that the Republican Party in the US use increasing cultural cleavages to attract voters with strong cultural preferences, while changing focus to economic issues after the election.[15] In my model, a coalition occurs without deception or changes in cultural preferences. Roemer (1998) constructs a model with strategic choices in two dimensions, and shows that the parties may converge to the median preference on the cultural dimension if cultural issues are more salient. Lee and Roemer (2006) calibrate such a model finding that racism makes *both* major parties in the US propose lower tax rates.[16] My approach is different, as I want to understand divergence in two dimensions. To the best of my knowledge, this paper is the first paper that links the development of polarization and right-wing populism with increasing inequality through policy-motivated parties. This paper also builds on a large literature trying to explain the support of far-right parties. In political science and sociology this literature is reviewed in Golder (2016) and Rydgren (2007).

A well-known problem of modeling political competition in more than one dimension is that a pure equilibrium may fail to exist, and that any policy outcome is unstable (McKelvey, 1976; Plott, 1967).[17] The related multi-dimensional literature has used

[12]See also Acemoglu et al. (2013) and Polborn and Snyder (2017) for other models of respectively populism and polarization.

[13]See also Austen-Smith and Wallerstein (2006), Anesi and Donder (2009) and Lindqvist and Östling (2013).

[14]Marx et al. (1975) used Irish immigrants workers to England as an example of this mechanism.

[15]Bartels et al. (2006) provides a critique of this argument.

[16]Roemer and Van der Straeten (2005) calibrate a model for France.

[17]See Duggan (2005) and De Donder and Gallego (2017) for summaries of the literature on multi-dimensional electoral competition. Even a mixed Nash equilibrium may fail to exist because of

different approaches to circumvent this problem. One solution is to fix the position on one of the two policy dimensions, such that the choice variable is one-dimensional.[18] The probabilistic voting model offers another solution by smoothing the payoff functions (Lindbeck and Weibull, 1987). I argue that observed features of the two-dimensional policy competition may allow for a simple model structure. One underlying reason for why an equilibrium generally fails to exist in two-dimensional policy models is that a party at any given point in time can make any small change to its policy choices. However, certain cultural policy dimensions tend to be perceived as binary. For example, in a survey conducted by Gallup in June 2019, 52% of respondents identify as pro-choice and 43% identify as pro-life, while only 4% take a mixed position.[19] Other cultural cleavages such as elite vs anti-elite, urban vs rural, "Somewheres" vs "Anywheres" and nationalism vs globalism may also be interpreted as binary policy dimensions. On the other hand, it is not common that people either support a tax rate of zero or a tax rate of one. Hence, I allow the policy in the economic dimension to be continuous, while the cultural policy space is binary. A related argument is that the cultural positioning may be perceived as more fundamental. It is probably easier for a politician to change the proposed tax rate from $0.3$ to $0.29$ than to change view on abortion or gun control. I model the game sequentially, such that the cultural policy is decided before the economic policy. The sequentiality of the game and the discreteness of the cultural dimension means that there are four one-dimensional subgames in the second stage of this game. The source of uncertainty in this model is related to the share of voters preferring each cultural policy. This implies that the median voter model by Meltzer and Richard (1981) occurs as a special case of this model conditional on the parties choosing the

---

the discontinuity of payoffs. Duggan and Jackson (2005) outline conditions under which a mixed equilibrium exists.

[18]Dziubiński and Roy (2011), Krasa and Polborn (2012), Krasa and Polborn (2014), Egorov (2015) and Matakos and Xefteris (2017) take this approach. Xefteris (2017) finds an equilibrium when candidates can influence $n - k$ dimensions assuming candidates are sufficiently differentiated on the other $k$ dimensions.

[19]https://news.gallup.com/poll/1576/abortion.aspx

same cultural policy.

The predictions from my model are supported by empirical evidence. Tavits and Potter (2015) find that more inequality is associated with right-wing parties emphasizing values-based issues, while left parties focus more on economic issues when there is a higher level of inequality. De La O and Rodden (2008) argue that secular and rich voters are more likely to vote according to their economic preferences, while the religious and poor voters are more likely to vote according to their religious preferences.

This paper is structured as follows. In Section 1.2 the main model is described, and this model is analyzed in Section 1.3. Section 1.4 presents extensions of this model, while Section 1.5 shows that the results are robust to changes in the set-up of the model. Section 1.6 provides an analysis of a simplified model with exogenous tax preferences.

## 1.2 The model

**Parties** I construct a model with two parties, $j \in \{r, l\}$, with mixed motivations; the parties want to win the election and implement tax rates close to their bliss points.[20] The probability that party $j$ wins the election is given by $P_j(\cdot)$, while $U_j(t_j)$ is the utility for party $j$ from implementing the tax rate $t_j$. The payoff function of party $j$ is given by

$$P_j(\cdot)U_j(t_j). \tag{1.1}$$

The right party, $r$, prefers a low tax rate, and the left party, $l$, prefers a high tax rate. For convenience I let the utility functions be linear in the tax rate and given

[20] In Section 1.5.1, I analyze parties that are purely policy-motivated.

by $U_r(t_j) = 1 - t_j$ and $U_l(t_j) = t_j$. The parties first simultaneously propose cultural policies, $c_j \in \{0, 1\}$, and conditional on the observed cultural policies, the two parties simultaneously propose tax rates given by $t_j \in [0, t_{max}]$. To keep the language simple, I will refer to $c = 1$ as a populist cultural policy. More precisely, $c = 1$ may refer to any policy that is preferred by a minority of voters. In line with traditional median voter models of redistribution, such as Meltzer and Richard (1981), I only allow for linear tax rates. If a party is indifferent between different tax rates, I assume that the party proposes its preferred feasible tax rate.

**Voters** The voters have different levels of productivity given by $\theta_i^\beta$, where $\theta_i \sim U[0, 2]$. The parameter $\beta$ measures the variation in productivity. I assume there is a competitive labor market, such that the wage rate given by $w(\theta_i)$ equals marginal productivity. The revenues from the flat tax rate $t$ is given by $T(t)$ and this amount is distributed evenly among the voters. The post-tax income of a voter is $(1-t)w(\theta_i)e + T(t)$, where $e$ denotes effort (e.g., hours worked), which has a convex cost given by $e^2$.[21]

The voters have preferences over the cultural policy. A share $1 - \alpha$ prefer $c = 0$, where $\alpha$ is distributed according to $\alpha \sim U[0, \frac{1}{2}]$. This assumption captures the fact that a party does not know the exact size of each group of voters, but knows that a majority of the voters always prefer $c = 0$.[22] This majority of voters get a negative payoff of $\delta$ if a party implements $c = 1$. Including the cultural policy, the indirect utility function for a voter with productivity $\theta_i$, belonging to the majority group, is given by

[21] In this model there is no income effect, which means that lowering the tax rate always increases the labor supply.

[22] I relax this assumption in Section 1.5.3.

$$U(t_j, c_j, \theta_i) = (1 - t_j)w(\theta_i)e(\theta_i, t_j) + T(t_j) - e(\theta_i, t_j)^2 - \delta c_j. \quad (1.2)$$

An indifferent voter is assumed to vote for each party with equal probability. The voters with populist cultural preferences get a payoff of $\delta_\alpha$ from $c = 1$. I want to focus the analysis on the majority group with non-populist preferences, so I assume that $\delta_\alpha$ is so large that the voters preferring a populist cultural policy always vote according to their cultural preferences.[23] This assumption makes the calculations easier, but it can be relaxed without changing the results. In Section 1.6, I show that the results are similar when the payoff from getting the preferred cultural policy is equal for voters that want to implement $c = 0$ and $c = 1$.

**Restrictions on parameter values and variables** The outcome of the model depends on the parameters $\beta$ and $\delta$. I solve the model under the assumption that $\beta \in [1, 5]$, given that a $\beta$ outside this interval does not match empirical wage data.[24] The majority of voters get a negative payoff $\delta$ by $c_j = 1$. When $\delta \to 0$, the cultural policy is unimportant for the majority of voters, and hence the model may provide multiple equilibria when $\delta$ is small. I limit the attention to cases where the equilibrium is unique, which occurs when $\delta > \frac{1}{6}$, and restrict $t_{max}$ to be the highest tax rate preferred by any voter. I also assume that party $r$ (party $l$) does not propose a tax rate that is strictly larger (smaller) than the preferred tax rate of the median voter.

**Timing of the game** The timing of the game is given by the following sequence of events.

[23]This is the case when $\delta_\alpha > 2^{2\beta}\frac{6\beta+1}{32\beta+16}$.

[24]Setting $\beta = 0$ e.g. means that all voters have the same productivity. The upper bound is not restrictive. When $\beta = 5$ the Gini coefficient is $\frac{5}{7}$, which is higher than the observed Gini coefficient for any country according to data from the World Bank: https://data.worldbank.org/indicator/si.pov.gini

1. The two parties, $j \in \{l, r\}$, simultaneously propose $c_j \in \{0, 1\}$.

2. For a given observation of $c_l$ and $c_r$, the parties simultaneously propose tax rates $t_j$.

3. Voters of type $\theta_i$ vote sincerely for one of the two parties.

4. The size of $\alpha$ is realized. The party that gets a majority of votes, $j^\star$, implements its proposed policy $\{t_j^\star, c_j^\star\}$.

5. The voters exert effort $e^\star(\theta_i, t_j^\star)$ and are paid a wage rate $w(\theta_i)$.

6. The collected tax revenues $T(t_j^\star)$ are evenly redistributed between all voters.

## 1.3 Analyzing the model

### 1.3.1 Subgame perfect equilibrium

I solve the model using backward induction. The equilibrium concept employed is subgame perfect Nash equilibrium, excluding weakly dominated strategies. Let $P_l(t_r, t_l, c_r, c_l)$ be the probability that party $l$ gets a majority of votes as a function of the different tax rates and cultural policies. The equilibrium in the second stage is given by

$$t_r^\star(c_r, c_l) = \arg\max_{t_r} [1 - P_l(t_r, t_l^\star(c_r, c_l), c_r, c_l)](1 - t_r),$$

$$t_l^\star(c_r, c_l) = \arg\max_{t_l} P_l(t_r^\star(c_r, c_l), t_l, c_r, c_l) t_l.$$

In the first stage, the equilibrium is given by

$$c_r^\star = \arg\max_{c_r}[1 - P_l[t_r^\star(c_r, c_l^\star), t_l^\star(c_r, c_l^\star), c_r, c_l^\star]][1 - t_r^\star(c_r, c_l^\star)],$$

$$c_l^\star = \arg\max_{c_l} P_l[t_r^\star(c_r^\star, c_l), t_l^\star(c_r^\star, c_l), c_r^\star, c_l])[t_l^\star(c_r^\star, c_l)].$$

### 1.3.2 Solution using backward induction

#### The wage structure

I have assumed that the conditions for a competitive labor market are satisfied, which implies that $w_i = \theta_i^\beta$. A larger $\beta$ can then be interpreted as more inequality, and the Appendix shows that typical inequality measures (such as the Gini coefficient, the Palma ratio, the 20:20 ratio and the coefficient of variation) are increasing in $\beta$.

#### The voters' choice of effort

The voters choose the optimal level of effort after the outcome of the election is realized. For a given tax rate $t$, the voters will choose effort to maximize

$$(1 - t)\theta_i^\beta e + T(t, \beta) - e^2.$$

The optimal effort is

$$e^\star(\theta_i, t) = \frac{(1 - t)\theta_i^\beta}{2}.$$

Intuitively, a larger tax rate discourages effort, while voters that are paid more will choose to exert more effort. The voters know the effort choices and wages of other voters, so they can calculate the tax revenues as a function of the tax rate. For a given tax rate $t$, the collected revenues per voter is $T(t) = t\int_0^2 \theta^{\beta}[\frac{e^{\star}(\theta,\beta,t)}{2}]f(\theta)d\theta = \frac{t(1-t)2^{2\beta}}{4\beta+2}$. The indirect utility function for a given policy combination $\{t_j, c_j\}$ for a voter with non-populist cultural preferences is therefore

$$U(t_j, c_j, \theta_i) = \frac{(1-t_j)^2\theta_i^{2\beta}}{4} + \frac{t_j(1-t_j)2^{2\beta}}{4\beta+2} - \delta c_j. \tag{1.3}$$

### 1.3.3 The voters' electoral choices

A voter with productivity $\theta_i$ and non-populist cultural preferences will vote for party $l$ if $U(t_l, c_l, \theta_i) > U(t_r, c_r, \theta_i)$. If $t_r = t_l$ and $c_r = c_l$, these voters are indifferent and vote for each party with equal probability. If $t_r = t_l$ and $c_r \neq c_l$, these voters will vote for the party proposing $c_j = 0$. If $t_r \neq t_l$, the share of voters with non-populist cultural policies voting for the left party is given by

$$s_l(t_r, t_l, c_r, c_l) = \frac{1}{2}\left(\frac{4\delta c_r - 4\delta c_l + \frac{2^{2\beta+2}}{4\beta+2}[t_l(1-t_l) - t_r(1-t_r)]}{(1-t_r)^2 - (1-t_l)^2}\right)^{\frac{1}{2\beta}}. \tag{1.4}$$

Equation (1.4) is defined when $s_l(\cdot) \in [0,1]$.[25] A voter belonging to the group with populist cultural preferences will vote for party $r$ ($l$) if $c_r > c_l$ ($c_r < c_l$), and otherwise vote according to the preferences over redistribution.

### 1.3.4 The choice of tax rates

There are four different subgames in the second stage depending on the different cultural policies chosen in the first period, and I can find the equilibrium in each of

[25] The share is more generally given by $s_{l''} = max[s_{l'}, 0]$, where $s_{l'} = min[1, s_l(\cdot)]$. I write $s_l(\cdot)$ to simplify the notation, and then I later check whether $s_l(\cdot) \in [0,1]$ is satisfied.

these subgames.

**Optimal tax rates when $c_l = c_r$**

**Lemma 1.** *If both parties propose the same cultural policy ($c_l = c_r$), then both parties will propose the median voter's preferred tax rate ($t_l = t_r = t^m(\beta)$).*

Lemma 1 and all following lemmas and propositions are proved in the Appendix. Although the parties are policy-motivated, convergence to the median voter's preferred tax rate is the unique equilibrium in these subgames. The median voter's preferred tax rate is found by maximizing equation (1.3) given $\theta_i = 1$, yielding

$$t^m(\beta) = \frac{2^{2\beta} - (2\beta + 1)}{2^{2\beta+1} - (2\beta + 1)}. \tag{1.5}$$

Suppose there is an equilibrium where the two parties propose the same cultural policy and different tax rates. If both parties win with positive probabilities (which occurs when the median voter is indifferent), then any party can make a small deviation towards the median tax rate and win with certainty. If one party wins with certainty, the other party wants to deviate to a tax rate closer to the median voter's preferred tax rate and get a positive payoff. Iterating this process the only equilibrium is $t_l = t_r = t^m(\beta)$. In the median solution, each party wins with probability $\frac{1}{2}$ while proposing the median tax rate.

**Optimal tax rates when $c_r = 1$, $c_l = 0$**

**Lemma 2.** *The unique equilibrium in the subgame where $c_l = 0$ and $c_r = 1$ is given by ($t_l = \frac{1}{2}, t_r = 0$).*

The proof consists of two steps. First I show that $t_r = 0$ is the best response to any feasible $t_l$. This simplifies the maximization problem of the left party, and I then show that $t_l = \frac{1}{2}$ is the best response to $t_r = 0$. In the Appendix, I prove Lemma 2 both analytically and numerically. For given values of $\beta$ and $\delta$, I find the equilibrium analytically, while a simulation-based approach is used to show that the proposed equilibrium is unique for *all* feasible values of $\beta$ and $\delta$. In this approach, I draw random values of $t_l$, $\beta$ and $\delta$ from their feasible regions. For all random draws I show that the best response of the right party is to propose $t_r = 0$. Given $t_r = 0$, I draw random values of $\beta$ and $\delta$ to show that $t_l = \frac{1}{2}$ is the best response for party $l$ to $t_r = 0$.

The two parties will propose their preferred tax rates in this subgame. A tax rate of $\frac{1}{2}$ maximizes tax revenues and is hence preferred by the left party and the poorest voters. The intuition behind Lemma 2 is that the difference in the cultural policy dimension will decrease the benefits of approaching the median tax rate. The marginal gain in votes is not equally large for the left party when moving from $t_l = \frac{1}{2}$ in this subgame. There are two groups of voters *not* voting for the left party. One group consists of the voters preferring a populist cultural policy, and a marginal decrease in $t_l$ will not attract any of these voters. The other group consists of the richest voters. A small decrease in $t_l$ will not attract many of these voters because of the skewness of the income distribution.

**Optimal tax rates when $c_r = 0, c_l = 1$**

**Lemma 3.** *The unique equilibrium in this subgame is for the right party to propose $t_r = \tilde{t}(\beta, \delta) > 0$ and win with probability $1$.*

Lemma 3 is proved using a similar approach to the approach from Lemma 2. The equilibrium in this subgame differs from the equilibrium in the subgame where $c_r = 1$ and $c_l = 0$. The party that does not propose a populist cultural policy faces a trade-off in all subgames where $c_r \neq c_l$. This party can propose its favorite tax rate or propose a tax rate closer to the median tax rate to increase its winning probability. In this subgame, the right party can win the election with certainty by making a small change in the tax rate away from its bliss point. The left party needs to make a larger deviation from its preferred tax rate to win the election with certainty in the subgame where $c_r = 1$ and $c_l = 0$. Intuitively, one can observe that the voters' utility functions are not symmetric around their bliss points. The poorest voter prefers $t = \frac{1}{2}$ and the richest voter prefers $t = 0$. But the rich voter faces a much larger drop in utility from a small change in the tax rate away from the bliss point, which means that the left party needs to propose a low tax rate to receive the support of rich voters. The derivative of the indirect utility function with respect to the tax rate is given by

$$\frac{\partial U(t_j, c_j, \theta_i)}{\partial t_j} = -\frac{(1-t_j)\theta_i^{2\beta}}{2} + \frac{2^{2\beta}}{4\beta+2}(1-2t_j). \tag{1.6}$$

I evaluate this derivative for the poorest and richest voters in their respective bliss points, which leads to $\frac{\partial U(t_j=\frac{1}{2}, c_j, \theta_i=0)}{\partial t_j} = 0$ and $\frac{\partial U(t_j=0, c_j, \theta_i=2)}{\partial t_j} = -2^{2\beta}\frac{2\beta}{4\beta+2}$.

Poor voters want a high tax rate, but they also know that a higher tax rate leads to less effort. A small reduction in the tax rate from $t = \frac{1}{2}$ is not optimal, but it leads to higher effort by all voters. For the richest voters a small increase in the tax rate is harmful for two reasons. For given effort levels the rich voters prefer a low tax rate, but a small increase in the tax rate additionally implies that all other voters provide less effort.

I here provide an example when the parameter values are given by $\beta = 2$ and $\delta = \frac{1}{4}$. In the subgame where $c_l > c_r$, the right party wins with certainty when proposing

$t \geq \tilde{t}(2, \frac{1}{4}) \approx 0.1$. On the other hand, in the subgame where $c_r > c_l$, the left party must propose a tax rate lower than $0.04$ to secure electoral victory. This example shows that the right party can guarantee electoral victory by making a small deviation away from its bliss point if $c_r < c_l$, while the left party needs to propose a tax rate close to zero to guarantee electoral victory if $c_r > c_l$.

### 1.3.5 First stage solution

In the first stage, the game is reduced to a binary game, where the two parties propose cultural policies. I show that $c_l = 1$ is a dominated strategy for the left party. If the right party chooses $c_r = 0$, the outcome will be given by the median voter's preferred tax rate, $t^m(\beta)$, and equal winning probabilities. This leads to a payoff of $\frac{1-t^m(\beta)}{2} = \frac{2^{2\beta-1}}{2^{2\beta+1}-2\beta-1}$ for the right party. If the right party chooses $c_r = 1$ in the first stage, the tax rates will be given by $t_r = 0$ and $t_l = \frac{1}{2}$. Equation (1.7) shows the payoff for the right party from choosing $c_r = 1$.[26]

$$P_r(\beta, \delta) = \frac{2}{[\frac{4}{3}(\frac{2^{2\beta}}{4\beta+2} + 4\delta)]^{\frac{1}{2\beta}}} - 1. \tag{1.7}$$

**Proposition 1.** *The unique subgame perfect equilibrium is given by* $\{t_r^\star(c_r^\star, c_l^\star) = 0, c_r^\star = 1\}$ *and* $\{t_l^\star(c_r^\star, c_l^\star) = \frac{1}{2}, c_l^\star = 0\}$ *if* $P_r(\beta, \delta) \geq \frac{1-t^m(\beta)}{2}$. *Otherwise, the unique subgame perfect equilibrium is convergence to the median voter's preferences given by* $\{t_r^\star(c_r^\star, c_l^\star) = t^m(\beta), c_r^\star = 0\}$ *and* $\{t_l^\star(c_r^\star, c_l^\star) = t^m(\beta), c_l^\star = 0\}$.

The equilibrium outcome depends on inequality through several different channels. The median voter prefers a higher tax rate when there is more inequality. This means that more inequality makes the median equilibrium less attractable for the right party.

[26]For notational simplicity, I write the winning probability for the right party in this subgame as $P_r(\beta, \delta)$ rather than $P_r[t_r^\star = 0, t_l^\star = \frac{1}{2}, c_r^\star = 1, c_l^\star = 0, \beta, \delta]$

The payoff when proposing a populist cultural policy is a non-monotonic function of inequality. The intuition is that more inequality makes the rich voters more willing to accept a populist cultural policy in order to implement their preferred tax rates. However, the set of rich voters is also becoming smaller when there is very much inequality.

### 1.3.6 Comparative statics

**Effect of inequality on cultural polarization**

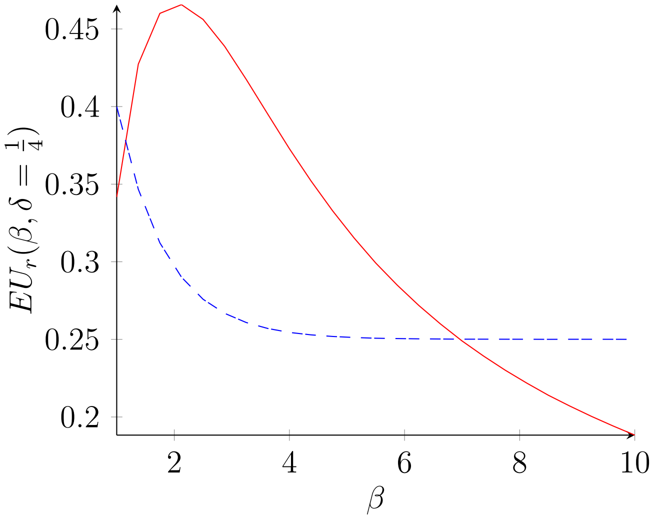
The right party proposes a populist cultural policy if $y(\beta,\delta) = P_r(\beta,\delta) - \frac{1}{2}[1 - t^m(\beta)] \geq 0$, and here I analyze how different levels of inequality ($\beta$) affects the right party's incentives.

**Lemma 4.** *An increase in inequality increases the incentives to propose $c_r = 1$ if $\frac{\partial y(\beta,\delta)}{\partial\beta} = \frac{\partial P_r(\beta,\delta)}{\partial\beta} + \frac{\partial t^m(\beta)}{2\partial\beta} \geq 0$, where:*
*i) $\frac{\partial t^m(\beta)}{\partial\beta} > 0$,*
*ii) $\frac{\partial P_r(\beta,\delta)}{\partial\beta} \geq 0$ if $\beta$ is smaller than some threshold.*

Increasing inequality has two different effects for the right party. More inequality leads to a higher median tax rate ($\frac{\partial t^m(\beta)}{\partial\beta} > 0$). This higher tax rate decreases the payoff for the right party in the median equilibrium, which increases the incentives to propose $c_r = 1$.

But increasing inequality also affects the share of voters preferring $c_r = 1$, and the effect of inequality on $P_r(\beta,\delta)$ is ambiguous. More inequality makes the richer voters more willing to choose a populist cultural policy in order to get their preferred tax rate, but it also makes the set of rich voters smaller. For a very large level of inequality, the set of rich voters becomes too small to be beneficial for electoral

**Figure 1.1:** Payoff for the right party when $c_r = 0$ (dashed line) and $c_r = 1$ (solid line).




purposes. This implies that the winning probability is increasing in $\beta$ only when $\beta$ is not too large, and that $P_r(\beta, \delta)$ gets very small when $\beta$ gets very large.

I fix $\delta = \frac{1}{4}$ to graphically describe the effect of inequality on the electoral outcome, and this is shown in Figure 1.1. This figure shows the expected payoff for the right party for the two choices of cultural policy as a function of inequality. The payoff in the median equilibrium (the dashed line) is decreasing in $\beta$, but is bounded below by $\frac{1}{4}$.[27] The payoff when proposing a populist cultural policy (the solid line) is a non-monotonic function of inequality. A populist cultural policy is more tempting for intermediate levels of inequality, and Proposition 1 shows that this result also holds for other values of $\delta$. When $\beta$ gets very large, the share of rich voters becomes electorally negligible, such that the right party does not propose a populist cultural policy.

[27] The tax revenues, $T(t)$, are maximized when $t = \frac{1}{2}$, so even when the level of inequality is very high, the median voter does not want to set $t > \frac{1}{2}$. This means that the right party gets at least a payoff of $\frac{1-\frac{1}{2}}{2} = \frac{1}{4}$ in the median equilibrium.

**Effect of inequality on the left party and the expected tax rate**

In a two-dimensional game the right party extends the electoral competition to a second dimension to be able to propose a lower tax rate. This polarization also enables the left party to propose a higher tax rate. The effect on the expected implemented tax rate is potentially ambiguous as one party proposes a lower tax rate and the other party proposes a higher tax rate, but the expected tax rate is lower in the polarized equilibrium for almost all parameter values. The intuition is that the right party chooses a populist cultural policy to be able to win the election with a large probability while proposing a low tax rate. The right party will not make this deviation if the winning probability is sufficiently small, and hence the right party will not deviate if the expected tax rate is small. When the right party proposes a populist cultural policy, the expected tax rate is

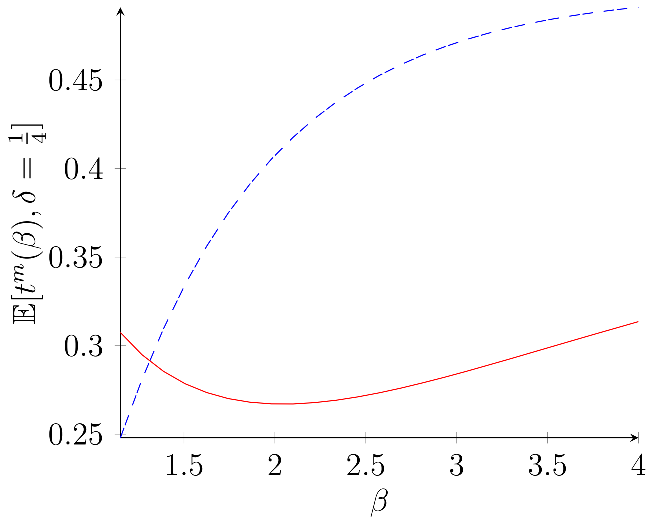
$$\mathbb{E}[t_{c_r=1}(\beta,\delta)] = P_r(\beta,\delta)t_r + (1-P_r(\beta,\delta))t_l = 1 - \frac{1}{[\frac{4}{3}(\frac{2^{2\beta}}{4\beta+2}+4\delta)]^{\frac{1}{2\beta}}}. \quad (1.8)$$

**Proposition 2.** *The expected tax rate in the polarized equilibrium is lower than the preferred tax rate of the median voter for all values of $\delta$ if $\beta > \frac{3}{2}$. If $\beta \leq \frac{3}{2}$, the tax rate in the polarized equilibrium is lower than the preferred tax rate of the median voter as long as $\delta$ is small.*

In a one-dimensional electoral game, more economic inequality is associated with a higher tax rate, which is described by Equation (1.5). Following divergence in two dimensions, Equation (1.8) shows a non-monotonic relation between inequality and taxation. Figure 1.2 compares the expected tax rate in the polarized equilibrium (the solid line) with the median tax rate (the dashed line) for $\delta = \frac{1}{4}$.[28]

[28] De Mello and Tiongson (2006) argue that the relation between inequality and redistribution is U-shaped.

**Figure 1.2:** Expected tax rate when the right party proposes $c_r = 0$ (dashed line) and $c_r = 1$ (solid line).




## 1.4 Extensions

### 1.4.1 Three parties

So far I have analyzed the electoral competition with two parties. This is a reasonable assumption in some countries (e.g., the US), while other countries are characterized by multi-party systems.[29] In this section, a third party is introduced. In addition to party $l$ and party $r$, there is a populist party $p$, and I want to analyze the incentives for engaging in different coalitions. The populist party is restricted to proposing $c_p = 1$, while parties $l$ and $r$ only propose a policy in the economic dimension ($t \in [0, t_{max}]$). The populist party wants to win the election while proposing $c_p = 1$, while the other parties have the same payoff functions as in previous sections.

#### Timeline of the game

1. The populist party proposes a coalition with party $j \in \{r, l\}$.

[29] Lijphart (2012) displays the number of effective parties for 36 different democracies.

2. $j$ accepts or declines the proposal.

3. If $j$ accepts, then a coalition between parties $j$ and $p$ proposes $c_p = 1$. The coalition and the non-aligned party then simultaneously propose tax rates. If $j$ declines the proposal, the populist party can propose a coalition with party $-j$.

4. $-j$ accepts or declines the proposal.

5. If $-j$ accepts, then a coalition between parties $-j$ and $p$ proposes $c_p = 1$. The coalition and the non-aligned party then simultaneously propose tax rates. If $-j$ declines the proposal, then parties $l$ and $r$ make a coalition (or alternatively, rule by minority government).

**Solution by backward induction**

If one of the parties accepts the coalition proposal, this coalition will propose $c = 1$ while the other party will propose $c = 0$. This means that the following stages in this version of the model are similar to the two subgames from Sections 1.3.4, where the two parties have chosen different cultural policies. Hence the left party will not accept the proposal, while the right party will accept the proposal if $P_r(\beta, \delta) \geq \frac{1-t^m(\beta)}{2}$.

**Proposition 3.** *The equilibrium is given by a coalition between the populist party and the right party proposing* $c_p = 1$ *and* $t_r = 0$ *if* $P_r(\beta, \delta) \geq \frac{1-t^m(\beta)}{2}$. *Otherwise, the left and right parties form a coalition or rule in a minority government.*

**Comparative statics**

$\delta$ is a measure of how much a populist cultural policy is disliked by the voters opposing $c = 1$, which means that $\frac{\partial P_r(\beta,\delta)}{\partial \delta} < 0$. If the populist party is too extreme (if $\delta$ is sufficiently large) the right party will not accept a coalition with the populist party. A comparison between Norway and Sweden may serve as an illustration of this effect. These countries have relatively similar electoral institutions, and both countries have populist parties given by *The Progress Party* (*FrP*) in Norway and *The Sweden Democrats* (*SD*) in Sweden. *The Progress Party* was previously isolated in Norwegian politics, and has successfully engaged in a process to become accepted by the other parties.[30] *The Progress Party* has later entered into a coalition government with the mainstream right-wing party in Norway, while *The Sweden Democrats* are politically isolated in Sweden.

## 1.5 Robustness

### 1.5.1 Policy-motivated parties

In the main model, I assume parties that care about the probability of winning the election and the implemented tax rate, but not the tax rate implemented by the other party. Here I analyze parties that are purely office-motivated, which means that the payoff function of party $j$ is given by $EU_j^{policy} = P_j(\cdot)U_j(t_j) + [1 - P_j(\cdot)]U_j(t_{-j})$. In such zero-sum games, there will often not exist a pure equilibrium, but making a small change in the structure of the game will lead to a pure equilibrium in this model. I assume that one (or none) of the parties is randomly drawn to be able to choose $c \in \{0, 1\}$ in the first stage, while the other party is restricted to propose $c = 0$. This assumption is primarily motivated by concerns for analytical tractability, but it is not necessarily unrealistic to assume that a party only in certain settings may be able strategically choose the policy on the second dimensions.[31] An indifferent party chooses the tax rate that maximizes the winning probability, and chooses the

[30] This process is described (in Norwegian) in this newspaper article https://www.klassekampen.no/59417/article/item/null/skal-gjore-frp-spiselig

[31] The cultural dimension is for example often modeled as fixed in the related literature.

tax rate closest to its bliss point if the winning probability is the same for all tax rates.

**Timing of the game**

1. Nature draws maximum one party, $j \in \{r, l\}$, to propose $c_j \in \{0, 1\}$.

2. The parties engage in the game described in Section 1.2, starting from stage 2 and with modified payoff functions given by $EU_j^{policy}$.

**Proposition 4.** *The unique subgame perfect equilibrium is given by* $t_r^\star = 0$, $c_r^\star = 1$ *and* $c_l^\star = 0$ *if the right party gets the possibility to choose a cultural policy and* $P_r(\beta, \delta) > 1 - 2t^m(\beta)$. *Otherwise, the unique subgame perfect equilibrium is given by convergence to* $t_l^\star = t_r^\star = t^m(\beta)$ *and* $c_l^\star = c_r^\star = 0$.

Depending on the level of inequality, introducing pure policy-motivation may make it more or less likely for a diverging equilibrium to occur. In the main model, there is divergence if $P_r(\beta, \delta) \geq \frac{1-t^m(\beta)}{2}$. When $t^m(\beta) \geq \frac{1}{3}$, which occurs when $\beta \geq \frac{3}{2}$, then pure policy-motivation makes the right party *more* willing to propose $c_r = 1$. When there is a high level of inequality, the tax rate preferred by the median voter is more similar to the high tax rate preferred by the left party. The polarization in tax rates induced by the cultural polarization will then be less important.

### 1.5.2 Diminishing marginal utility

The voters' marginal utility of income is constant in the main model. Alternatively, I can allow for diminishing marginal utility of income, given by an isoelastic utility function with parameter $\rho \geq 0$, which implies that $u(w) = \frac{w^{1-\rho}}{1-\rho}$. The equilibrium from the main model is generally not a knife-edge equilibrium, which means that a small increase from $\rho = 0$ should not affect the equilibrium outcome. In this section, I provide an informal discussion of how $\rho > 0$ may affect the incentives to propose a populist cultural policy, while a formal analysis is given in the Appendix.[32]

Diminishing marginal utility of income will affect the voters' effort levels, and the effect may be heterogeneous. Less productive voters may increase the effort level when $\rho > 0$, while more productive voters decrease their level of effort. This implies that the effect of $\rho > 0$ on the collected tax revenues is ambiguous. This also implies that the effect of $\rho > 0$ on the median voter's preferred tax rate is ambiguous.

The main model shows that the right party proposes a populist cultural policy if the probability of winning the election while bundling this policy with a low tax rate is large relative to the median voter's preferred tax rate. One effect of $\rho > 0$ may be that rich voters care less about an additional unit of income, and the isolated effect is that it becomes less likely for the right party to win while proposing a different cultural policy. But $\rho > 0$ also has another effect on the winning probability while proposing $c_r = 1$. The rich voters compare the payoff from $c_r = 1$ and $t_r = 0$ with the payoff in the median equilibrium, and this latter payoff may also decrease in $\rho$ if a larger $\rho$ increases the median tax rate. A higher median tax rate does not only affect the electoral choices of rich voters, but also affects the payoff for the parties in the median equilibrium. If $\rho > 0$ increases the median tax rate, then the right party may be more willing to accept a lower winning probability in order to implement its preferred tax rate.

[32]I will not provide a characterization of the equilibrium given diminishing marginal utility of income, bur rather show that the equilibrium effects may be ambiguous.

### 1.5.3 More voters preferring $c = 1$

Cultural polarization is chosen in the main model to enable different tax rates in equilibrium, and I have shown that the right party is more willing to propose a populist cultural policy. When there is a high level of inequality, the right party is less satisfied with the median tax rate, and may engage in cultural polarization although this lowers the winning probability. Turning around the argument, the left party does not want to propose a populist cultural policy because this party is more satisfied with the median tax rate, and is not willing to accept a lower probability of winning to get an even more beneficial tax rate. In this section, I analyze the incentives to propose $c = 1$ when such a policy is preferred by more voters. I let the share of voters preferring $c = 1$ be drawn according to $\alpha \sim U[0, \alpha_{max}]$, where $\alpha_{max} \geq \frac{1}{2}$.

**Proposition 5.** *There is an equilibrium where $c_l^\star = 1$ and $c_r^\star = 0$ if $\alpha_{max} \geq \frac{1}{2[1-t^m(\beta)]}$. In this equilibrium, the right party proposes $t_r^\star = \tilde{t}(\beta, \delta) < t^m(\beta)$, while the left party proposes $t_l^\star = \frac{1}{2}$.*

When $\alpha_{max}$ is large relative to $t^m(\beta)$, there exists an equilibrium where the left party proposes a populist cultural policy. As in previous sections, the cultural polarization enables the two parties to diverge from the median tax rate in equilibrium. More inequality (larger $\beta$) makes the left party less willing to propose $c_l = 1$. More inequality means that the payoff for the left party in the median equilibrium increases, which reduces the incentives to diverge from the median equilibrium. Another insight is that $\alpha_{max} = \frac{1}{2}$ is not the threshold for the uniqueness of the equilibrium in the main model. Even when inequality is low, $\alpha_{max} \geq \frac{5}{8}$ is a necessary condition for the left party to propose $c_l = 1$.

## 1.6 A model with exogenous tax preferences

The main model is embedded in a framework where the labor supply and the preferences for taxation arise endogenously. Here I construct a simpler model to show that the main results also holds outside of the framework from the main model. To make a model as simple as possible, I let the set of possible tax rates be binary and given by $t \in \{t_{low}, t_{high}\}$. The right party prefers a low tax rate and the left party prefers a high tax rate. A party gets a payoff of $U_p$ from winning the election and implementing its preferred tax rate, and a payoff of $U$ from winning the election and implementing the non-preferred tax rate. I construct the timing of the game to be similar to the main model.

#### Timing of the game

1. The two parties, $j \in \{r, l\}$, simultaneously propose $c_j \in \{0, 1\}$.
2. For a given observation of $c_l$ and $c_r$, the two parties simultaneously propose tax rates $t_j \in \{t_{low}, t_{high}\}$.
3. The size of $\alpha$ is realized and voting takes place.

As in the main model, a share $\alpha < \frac{1}{2}$ of voters prefer $c = 1$, and there is some uncertainty about the size of $\alpha$. Let $P_{1,0}$ denote the probability that party $r$ wins the election given $c_r = 1$, $c_l = 0$, $t_r = t_{low}$ and $t_l = t_{high}$, while $P_{0,1}$ denotes the probability that $r$ wins the election given $c_r = 0$, $c_l = 1$, $t_r = t_{low}$ and $t_l = t_{high}$. I assume that $0 < P_{1,0} < \frac{1}{2}$ and $0 < 1 - P_{0,1} < \frac{1}{2}$, which means that a party that proposes $c = 1$ cannot win the election with a probability greater than one half.

**Proposition 6.** *Suppose a majority of voters prefer $t_{high}$. The unique subgame perfect equilibrium is given by $c_r^\star = 1$, $c_l^\star = 0$, $t_r^\star = t_{low}$ and $t_l^\star = t_{high}$ if $\frac{U_p}{U} \geq$*

*$max[\frac{1}{1-P_{1,0}}, \frac{1}{2P_{1,0}}]$. Otherwise, the only pure subgame perfect equilibrium is given by $c_r^\star = c_l^\star = 0$ and $t_l^\star = t_r^\star = t_{high}$.*

In the main model, the right party has more incentives to engage in cultural polarization because the party dislikes a high level of inequality *and* because rich voters are more willing to make cultural compromises. Proposition 6 isolates the first of these two effects. When there is a high level of inequality, the interests of a majority of the voters are aligned with the left party. The right party is unable to win the election if the only difference between the parties is given by $t_r < t_l$. For the right party to be able to implement $t_{low}$ when a majority of voters prefer $t_{high}$, it is necessary that $c_r \neq c_l$. Bundling $t_r = t_{low}$ with $c_r = 1$ leads to a winning probability of $P_{1,0} < \frac{1}{2}$, but makes it possible to win the election while proposing a low tax rate. The above insight can be reversed when there is a *low* level of inequality. I now introduce the preferences of the voters to the above framework.

**Voters** Each voter is characterized by a wage $w_i$ and cultural preferences $\delta$. The wage is drawn from some cumulative distribution $G[\cdot]$. The cultural preferences are binary and the type is given by $\tau \in \{-1, 1\}$. An unknown share $\alpha$ of voters are of type $\tau = 1$ and get a positive payoff $\delta$ from $c = 1$, while a share $1 - \alpha$ get a negative payoff of $-\delta$ from $c = 1$. The total utility of a voter of type $\{w_i, \tau\}$ for a policy $\{t_j, c_j\}$ is given by

$$(1 - t_j)w_i + T(t_j) + \delta\tau c_j.$$

This set-up means that I can find expressions for $P_{1,0}$ and $P_{0,1}$. I now analyze the incentives of the parties to induce polarization when party wants a tax rate that is preferred by a minority of the voters. To isolate the second effect from the main model, I let the median voter be close to indifferent between $t_{high}$ and $t_{low}$. This means that I analyze the incentives of the parties to propose $c = 1$ and their preferred tax rate when the median voter *marginally* prefers the other tax rate.

**Proposition 7.** *Suppose the median voter is close to indifferent between $t_{high}$ and $t_{low}$. $P_{1,0} \geq 1 - P_{0,1}$ as long as $G[\cdot]$ is concave.*

Proposition 7 outlines the conditions such that the right party attracts a larger coalition of voters from engaging in cultural polarization. When the income distribution is concave, the coalition of voters that accepts a non-preferred cultural policy to implement a preferred tax rate is larger for the right party. A concave income distribution means that the income levels of the richest voters are further away from the median income. In this model, the richest voters are more willing to make compromises on the implemented cultural policy to get their preferred tax rate. Concavity of the income distribution arises in a large range of settings.[33]

## 1.7 Conclusion

The main contribution of this paper is to present an argument connecting the observed trends of right-wing populism and polarization with the observed relation between economic inequality and redistribution. As described in previous sections, there is a large literature aiming to explain the weak relation between economic inequality and redistribution, and the populism literature is also expanding. These topics are large and important. Naturally, I do not claim that the argument presented in this paper is

[33]The Appendix provides an histogram of the US income distribution, and this distribution approximately satisfies concavity. A Pareto (Type 1) distribution has a cdf given by $F(x) = 1 - [\frac{x_m}{x}]^{\gamma}$, where $x > x_m$ and $\gamma > 0$, and this function is concave. The cumulative distribution function of productivity in the main model of this paper is given by $G(w) = P(\theta^{\beta} < w) = \frac{w^{\frac{1}{\beta}}}{2}$, and this function is concave when $\beta > 1$, as the second derivative is given by $G''(w) = \frac{1-\beta}{2\beta^2} w^{\frac{1-2\beta}{\beta}}$.

the only explanation to these major trends. I rather want to identify one argument that may cast some light on these empirically observed patterns.

Another contribution of this paper is to propose a model structure that allows for equilibrium in multi-dimensional policy spaces. I restricted the analysis to two dimensions, where one dimension is binary. The framework can be extended to $n$ dimensions as long as $n-1$ dimensions are discrete, and such a model structure may provide a natural space for analyzing multi-dimensional games in subsequent research.

# Appendix A. Inequality metrics and income distributions

I interpreted an increase in $\beta$ as an increase in inequality, and this is satisfied when using the typical metrics for income inequality. The Gini coefficient ($G$) is defined according to

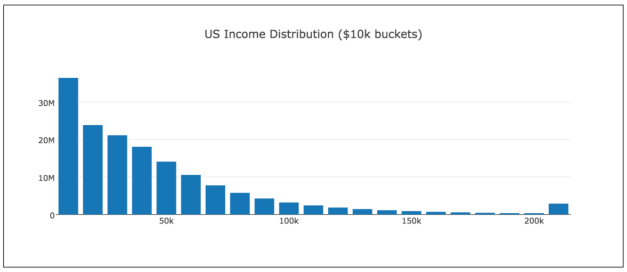
$$G = \frac{\int_0^\infty F(w)(1-F(w))dw}{\mu}.$$

In the main model, $\mu = \frac{2^\beta}{\beta+1}$ and $F(w) = \frac{w^{\frac{1}{\beta}}}{2}$, which implies that the Gini coefficient is given by

$$G = \frac{\beta+1}{2^\beta} \int_0^{2^\beta} \frac{w^{\frac{1}{\beta}}}{2}[1 - \frac{w^{\frac{1}{\beta}}}{2}] = \frac{\beta}{\beta+2}.$$

The Gini coefficient is clearly increasing in $\beta$. When $\beta = 1$ the Gini is given by $\frac{1}{3}$, while $\beta = 2$ corresponds to a Gini coefficient for productivity of $\frac{1}{2}$. A larger $\beta$ also leads to a larger Palma ratio and a larger 20:20 ratio. The 20:20 ratio is given by $\frac{2^{(\beta+1)}-1.6^{(\beta+1)}}{0.4^{(\beta+1)}}$ and the Palma ratio is given by $\frac{2^{(\beta+1)}-1.8^{(\beta+1)}}{0.8^{(\beta+1)}}$. Both of these are increasing in $\beta$. The variance is given by $2^{2\beta}[\frac{1}{2\beta+1} - \frac{1}{(\beta+1)^2}]$, which means that the coefficient of variation is given by $\frac{\beta}{\sqrt{2\beta+1}}$. This expression is also increasing in $\beta$.

**Income distributions** Income inequality typically implies a distribution that is positively skewed. In the US, one can typically observe income distribution like the picture below.[34]




This graph has uniform scaling. Each bar represents the people up with incomes in the $10k bucket range up to the current x-axis label. The last bucket is the number of people with incomes $200,000 and greater.

**Figure 1.3:** Income distribution in the US

Figure 1.3 shows that the empirical density of the income distribution is decreasing. If the derivative of the cumulative distribution of income is positive and decreasing it implies that the cumulative distribution function is concave.

# Appendix B. Diminishing marginal utility of income

In this part of the paper, I will try to repeat some of the previous analysis, but allowing for voters to have diminishing marginal utility of income. The voters are here characterized by an isoelastic utility function with parameter $\rho$. The purpose of this section is to show that diminishing marginal utility may have ambiguous effects, rather than to provide a characterization of the equilibrium. I simplify the utility function to be quasi-linear, which means that I am able to analyze how effort

[34]This figure uses US census data and is available from https://medium.com/jeremy-keeshin/which-percent-are-you-the-actual-income-distribution-in-the-united-states-1272d34b5b9b

depends on $\rho$.[35] This interpretation can e.g. be justified if the collected tax revenues $T(t,\rho)$ are spent on public goods. This is clearly a restrictive assumption, but the purpose of this section is to show that even when simplifying the utility function, the effects of $\rho > 0$ are complicated and unambiguous. The optimal effort maximizes

$$max \quad \frac{[(1-t)\theta_i^\beta e]^{1-\rho}}{1-\rho} + T(t,\rho) - e^2. \tag{1.9}$$

This leads to effort levels and tax revenues given by

$$e^\star(t,\theta_i,\rho) = \frac{1}{2^{\frac{1}{1+\rho}}}[(1-t)\theta_i^\beta]^{\frac{1-\rho}{1+\rho}},$$

$$T(t,\rho) = \frac{t}{2}\frac{1}{2^{\frac{1}{1+\rho}}}[(1-t)]^{\frac{1-\rho}{1+\rho}}\frac{1+\rho}{2\beta+1+\rho}2^{\frac{2\beta+1+\rho}{1+\rho}}.$$

I then analyze the effect of a small increase in $\rho$ from $\rho = 0$.

**Lemma 5.** *A marginal increase in $\rho$ from $\rho = 0$ will affect effort according to*

$$\frac{\partial e^\star(t,\theta_i,\rho)}{\partial\rho}\Big|_{\rho=0} = \frac{(1-t)\theta_i^\beta}{2}[\ln(2) - 2\ln[(1-t)\theta^\beta]].$$

*The effect on tax revenues is given by*

$$\frac{\partial T(t,\rho)}{\partial\rho}\Big|_{\rho=0} = T(t)[1 + 0.692(1-2\beta) - \frac{1}{2\beta+1} - 2\ln(1-t)]$$

The direction of both of these effects are ambiguous. Lemma 5 shows that a marginal increase in $\rho$ will increase effort if the voter has a low post-tax income, which

[35] If I do not assume the preferences to be quasi-linear, the first order condition for the optimal level of effort generally does not have an analytical solution.

happens if $(1-t)\theta^{\beta} < \sqrt{2}$. The effect on tax revenues is then also ambiguous. A small $\beta$ and a large $t$ implies that the derivative is positive.

Equation (1.9) shows that introducing $\rho > 0$ may have an ambiguous effect on the choices of the median voter, which means that the effect on the right party's incentives to propose $c_r = 1$ may also be ambiguous. The right party proposes a populist cultural policy if this leads to a payoff that exceeds the payoff in the median equilibrium. The payoff in each of these equilibria depends on how $\rho$ affects the median voter. As described in previous sections, the winning probability of the right party when proposing $c_r = 1$ depends on the effect of $\rho$ on $T(t, \rho)$, as this affects the size of the coalition of voters. Additionally, there is a direct effect of the median tax rate on the payoff for the right party when proposing $c_r = 0$.

# Chapter 2

# Population-Level Treatment Effects: Theory and Evidence from a Large-Scale Voting Experiment

Ole-Andreas Elvik Næss [1]

**Abstract**

This paper analyzes methods for drawing causal inference when a treatment is provided to an entire population rather than being randomly implemented. Causal inference is possible by randomizing the timing and order of the treatment. I send a text message with an encouragement to vote to all 200,000 voters in Bergen, Norway. By randomizing the timing and order of the text message, and using a novel dataset that includes the exact timing of each vote, I estimate that the text message increases the electoral turnout by 1.4-2 percentage points. I also argue that the underlying mechanism is related to provision of information (*JEL*: C93, D72) .

[1] I would like to seminar participants and discussants in Bergen and Oslo for helpful comments and discussions. I would like to extend a special thanks to Eirik Gaard Kristiansen and Tore Ellingsen for tremendous guidance and support. The usual disclaimer applies.

## 2.1 Introduction

Randomized controlled trials (RCTs) are often considered to be the gold standard of causal inference. However, a large share of population-level programs are provided to entire populations, rather than being randomly implemented. The reluctance to randomize, whether justified or not, raises the question of whether it is possible to conduct population-level studies that approach the quality of inference from RCTs.

The opposition to using RCTs to implement programs at the population level may be related to ethical or practical arguments, but some of this opposition may also be methodological.[2] For an RCT to provide an unbiased treatment effect, it is necessary that the non-receivers are unaffected by the treatment. This assumption is often innocuous in small experiments, but the assumption of no interference is stricter when a large share of the population is treated. An informational campaign or an encouragement to take a certain action may for example have spillover effects on non-receivers.

This paper makes two main contributions. First, I investigate how to evaluate causal treatment effects when we are constrained to providing the treatment to everyone in a given population. The reason behind this constraint may for example be related to practical, methodological or ethical issues. In the second part of the paper, I apply such methods to test the effect of a text message from the government encouraging people to vote, and I also analyze the underlying mechanisms behind the effect.

A central feature of statistical analysis is how to estimate population quantities from smaller samples. When all units in a population receive a treatment, there is no uncertainty about the value of aggregate data. However, there is still uncertainty

[2] The process of scaling up economic experiments has been analyzed by Al-Ubaydli et al. (2017), Muralidharan and Niehaus (2017) and Davis et al. (2017), while Deaton and Cartwright (2018) provide some critique of RCTs.

related to the outcome in the absence of a treatment. Abadie and Gardeazabal (2003) and Abadie et al. (2010) develop the synthetic control method to construct a comparison group for a treated aggregate unit in the absence of a treatment. Other methods may also be used to analyze population-level treatment effects when we have access to data that is precisely measured in time. Randomizing the *timing* of the treatment may provide a suitable counterfactual. Suppose a treatment is given to an entire population at a random point in time given by $t^\star$. If we make the assumption that the outcome variable is continuous in time, then the outcome at time $t^\star - s$ will be a proper counterfactual for the treated population in the absence of any treatment at time $t^\star + s$ when $s \to 0$. This is the logic underlying the regression discontinuity design.[3]

Even in cases where we are constrained to providing the treatment to all units, it may be possible - or even unavoidable - that the treatment is given to different units at marginally different points in time.[4] By randomizing the order of the treatment, we can get access to more treatment effects in short time periods around time $t^\star$.

I also argue that we can analyze the effect at $t^\star$ separately from the effect for longer periods of time. The estimated treatment effect for a longer period of time is biased if there are confounding variables, but such variables may not affect the outcome at time $t^\star$, because it is unlikely that the effect starts *exactly* at time $t^\star$. When estimating the treatment effect at time $t^\star$, we are concerned about something else happening around $t^\star$, but something happening in a small time interval will not have a large effect on a continuous outcome variable for a longer period of time.

[3] Lee and Lemieux (2010) provide an introduction to regression discontinuity design, while Hausman and Rapson (2018) show how to use regression discontinuity in time.

[4] Such variations may occur if the treatment is provided through phone calls or door-to-door canvassing.

### 2.1.1 A voting experiment

I want to understand if a government can increase the population-level electoral turnout using a nudging text message. There is a vast RCT literature encouraging individuals to vote, but to the best of my knowledge there are no studies on whether an encouragement will have the same effect if a large share of the population is treated.[5]

Low and unequal turnout is widely considered as a serious problem for democracies.[6] If a text message is able to increase the electoral turnout at the population level, then this tool may serve an important policy function. But it is not necessarily straightforward to assume that the treatment effect is similar when a large share of voters receive the same encouragement.[7]

A population-level RCT estimates the treatment effect of a text message as the difference in turnout between receivers and non-receivers given that a large share (for example fifty or eighty percent) of the voters receive a text message. In this setting, there may be both methodological and ethical concerns by using an RCT. There is ample evidence that voting encouragements affect the non-treated, which means that a population-level RCT may lead to a biased treatment effect. In particular, it has been shown that the effect of a voting encouragement is contagious within households (Nickerson, 2008; Sinclair et al., 2012) and that the effect also spreads through social networks (Bhatti et al., 2017a; Bond et al., 2012).[8] When the treatment

[5]The large individual-level literature is summarized in Green et al. (2013), Green and Gerber (2015) and Gerber and Green (2017). Voting encouragements are typically referred to as "Get Out The Vote" (GOTV) efforts.

[6]See, e.g., Verba et al. (1978), Wolfinger and Rosenstone (1980), Hill et al. (1995), Lijphart (1997), Lijphart (1998), Mahler (2008) and Fowler (2013).

[7]Green and Gerber (2015) for example find that encouragements that are perceived as personal communication have larger effects, and a text message sent to *everyone* is not a particular personal form of communication.

[8]Jones et al. (2017) find that the effect of an encouragement on close friends is larger than the direct treatment effect. More generally, a growing literature argues that political participation and voting must be understood as social decisions (DellaVigna et al., 2016; Gerber et al., 2008;

group is small relative to the population, it is arguably not a major problem that the effect is contagious within close relations. However, when sending a text message to a randomly drawn large share of the voters, then most of the non-treated voters will belong to the set of friends and family members of the treatment group. Another methodological issue relates to the fact that non-receivers are assigned to control by an authoritative figure (the government) with a clear goal of increasing turnout, which means that the setting is vulnerable to experimenter demand effects (Karakostas and Zizzo, 2016; De Quidt et al., 2017). This setting is particularly vulnerable to such effects because the expected treatment effect is small.[9]

There are also potential ethical issues with using an RCT in this setting. Randomizing the set of recipients may be perceived as interfering in the election.[10] Sending a text message to a random set of voters additionally means that researchers need access to individual voting records.[11]

Together with the local government in Bergen, Norway, I sent a text message to all voters in Bergen (199,918 voters) at a random point in time some days before the Parliamentary election in 2017. This text message contained an encouragement to vote in the upcoming election. Half of the districts in the city received the text message a few hours before the other districts. This setting is suitable for exploiting the time dimension of voting. Around $75,000$ voters in Bergen voted before the electoral day, and I have access to a unique data set with the *exact* timing of each Norwegian vote. I can also measure the timing of the treatment precisely, and some of the treatment effect may potentially be close to instantaneous.[12]

Perez-Truglia and Cruces, 2017).

[9] I later show that just a very small share of voters need to act according to their treatment status in order for this bias to be larger than the expected treatment effect.

[10] Although all voters received a text message, this project was later criticized in the local media based on the argument that a government should not interfere in an election. This suggests that only sending the text message to half of the voters would be a controversial strategy in this setting.

[11] Privacy concerns is not an issue when everyone is treated. In this case, identifying data is not needed as long as I collect information about the *timing* of each vote.

[12] There is a large number of places to vote early in Bergen, including libraries, universities, malls

**Results** I first establish that there was an unusually high electoral turnout in Bergen in 2017. I make comparisons across the 19 largest cities in Norway for the past 6 elections, and Bergen in 2017 shows the largest deviation from the average turnout of all 114 observations. Using the synthetic control method allows me to make comparisons with a more suitable counterfactual, which leads to estimated effects between 1.4 and 2.2 percentage points. I construct Placebo tests in space and time, and show that few or none of the 114 observations deviate as much from their synthesized versions as Bergen did in 2017.

These results show a high electoral turnout in Bergen in 2017, but they do not causally link the increased turnout to the text message. I use the randomization of the timing of the treatment to find out *when* the turnout increased in Bergen, and I find that the turnout in Bergen sharply increased after the voters received the text message. This immediate increase is so large that it is consistent with the text message being the cause of the high turnout in Bergen in 2017. The increase in turnout after the text message is the largest observed increase among all 20 days of early voting in Bergen. Also, I only observe this increase on the treatment day *after* the voters received the text message. The turnout on the early morning of the day of the treatment (before the text messages were sent out) is similar to the day before. The increase in turnout in Bergen is large compared to other cities, and the regression discontinuity (RD) estimate is large and statistically significant. I estimate that the text message increased turnout in Bergen on this particular day by approximately 1 percentage point. Furthermore, I also show that voting immediately increased among the first receivers relative to the group that had not yet received a text message. The RD estimate for the difference in voting between the two parts of the city is also large and statistically significant. There are no other days of early voting where there is such a large difference between the two parts of the city. Aggregating the difference

and IKEA.

from this short time period leads to an effect size that is consistent with the other treatment effects.

**Underlying mechanisms** I also investigate the underlying mechanisms. Two main arguments are presented in the literature for *why* an encouragement may be able to increase the electoral turnout. One argument states that encouragements work when they are personal and social. Green and Gerber (2015) argue that voters think of the encouragement as an invitation to a social event, and Gerber and Rogers (2009) suggest that descriptive social norms are important for explaining turnout. On the other hand, Dale and Strauss (2009) claim that an encouragement to vote works simply as provision of information, by acting as a noticeable reminder of the election. One half of the voters received a text message focusing on descriptive social norms, while the other half received a non-social content. I do not find any evidence indicating a different effect from the two messages. I present evidence indicating that the voters appreciated the informational content of the text message. I elicit the voters' opinions about the text message as a source of information by comparing two surveys. Some months prior to the election, a representative sample in Bergen ($n = 600$) were asked how they wanted the local government to provide information about important events, and $4\%$ wanted to be informed through text messages. A few months after the election, the survey asked another representative sample the same question, and in this case, $41\%$ of the recipients stated that they wanted the government to use text messages to provide information about elections. This change indicates that the mechanism driving the effect of the text message may be more related to provision of information than descriptive social norms.[13]

[13] I do not claim that information is more important than social norms in explaining turnout, but just that the effect of a text message (which is a non-social tool) may be more related to information than norms.

**Roadmap of the paper** This paper is structured as follows. In Section 2.2, I show how randomization in time will lead to an unbiased treatment effect. In Section 2.3, I describe the experimental design of the voting encouragement, while Section 2.4 presents the results from this experiment. This paper analyzes methods for causal inference when we are constrained to providing the treatment to everyone. Informally, I have argued that the opposition to using RCTs at the population level sometimes may be related to methodological concerns. In Section 2.5, I provide a formal analysis of this argument. I show that we need to make stronger assumptions to get unbiased treatment effects at the population level, and I also construct a model to analyze which parameter values that make population-level treatment effects most vulnerable to biases arising from interference.

## 2.2 Identification and estimation of unbiased treatment effects

In this section, I argue that we can use past outcomes as counterfactuals by randomizing the timing of the treatment and making comparisons in small time intervals around this point in time. The population-level outcome at time $t$ is given by $f_D(t)$, where $D \in \{0, 1\}$ denotes the binary treatment status. The causal effect of a treatment at time $t^\star$ between $t^\star$ and $t^\star + s$ is given by $\alpha(t^\star + s) = f_1(t^\star + s) - f_0(t^\star + s)$. The fundamental problem of causal inference (Holland, 1986) states that we cannot observe both $f_1(t^\star + s)$ and $f_0(t^\star + s)$.

**Assumption 1.** *$f_0(t)$ is continuous in time.*

When the outcome variable is continuous in time, the above assumption allows for comparisons around the threshold. This is similar to the identifying assumption underlying the regression discontinuity (RD) design, and Hausman and Rapson (2018) analyze the use of RD when time is the running variable.

### 2.2.1 Marginal variations in the timing of the treatment

I have argued that even in cases where we are constrained to providing the treatment to everyone, it may be possible to implement marginal variations in the timing of the treatment. The logic underlying the RD analysis also holds when the treatment is provided to different units at marginally different points in time.

I also construct a difference-in-difference estimator. If one group, $g = 1$, receives the treatment marginally before $g = 2$, I construct an estimator given by $\tilde{\alpha_1}^{DiD}(t^\star + s) = [f_1^{g=1}(t^\star + s) - f_0^{g=1}(t^\star - s)] - [f_0^{g=2}(t^\star + s) - f_0^{g=2}(t^\star - s)]$. In the unlikely case that something else happens at $t^\star$, it is unlikely that only group $g = 1$ will be affected.

### 2.2.2 Treatment effect for a longer period of time

I also want to analyze the total effect of the treatment, and not just the effect in a short time interval. We cannot use the continuity assumption to get causal effects for longer time periods, because we do not have a suitable counterfactual for the treated population in this case. Other methods may be used to estimate treatment effects. The synthetic control method (Abadie et al., 2010) uses other population units and observations from other time periods to construct a counterfactual.

### 2.2.3 Relation between effects from different time periods

When estimating the treatment effect for a longer period of time, a bias arises if another variable affects the outcome after the treatment. Here I argue that such a bias will often not influence the estimated effects in short time periods after $t^{\star}$.

Suppose another variable increases the outcome variable by some number $\mu$, starting from some point in time, given by $t'$. The probability that this occurs exactly at time $t^{\star}$ is close to $0$, given that the timing of the treatment is randomized.

The above argument assumes that the other variable has a constant treatment effect, but for certain classes of stochastic distributions, the argument also holds if the other variable is increasing in time. I now model the outcome variable, $f(t)$, as a Brownian Motion with drift, which is the only Lévy process with continuous paths. This process is given by

$$f(t+s) - f(t-s) \sim N(2\mu s, 2\sigma^2 s).$$

Here $\mu$ is the time trend of the other variable, $\sigma^2$ is the variance of the distribution and $G[\cdot]$ is the cumulative distribution function of the standard normal distribution. At time $t^{\star} - s$ the outcome variable is given by some value $f(t^{\star} - s)$.The probability that $f(t^{\star} + s)$ is larger than $cf(t^{\star} - s)$, where $c > 1$, is given by

$$P[f(t^{\star} + s) > cf(t^{\star} - s)] = 1 - G[\frac{(c-1)f(t^{\star} - s) - 2\mu s}{\sigma\sqrt{2s}}].$$

In a very short time interval after $t^{\star}$, the probability of observing a large change approaches zero, and we can verify that $\lim_{s \to 0} P[f(t^{\star} + s) > cf(t^{\star} - s)] = 0$.

**Proposition 8.** *The correlation between the change between time $t^{\star}$ and $t^{\star} + s$ and* *the change between $t^\star$ and $t^\star + \delta$, approaches zero when the outcome follows a Lévy process with continuous paths, and $s \to 0$ and $\delta > 0$.*

All proofs are provided in the Appendix. Proposition 8 is satisfied although there is an underlying time trend. We should not expect to find a correlation between changes in very short time periods and changes for longer periods of time when there is sufficient variation in the stochastic process.

## 2.3 A voting experiment

### 2.3.1 Experimental design

I cooperated with the local government of Bergen, Norway, in a project aiming to increase the electoral turnout in the 2017 Parliamentary Election. Recent experiments have shown that text messages may increase the electoral turnout (Bhatti et al., 2017b; Dale and Strauss, 2009). I wanted to find out if a nudging text message from the government can increase the electoral turnout at the population level. Nudges from the government have been shown to cost-effectively solve a range of policy problems (Benartzi et al., 2017; Sunstein and Thaler, 2008). A text message is a cheap intervention, so even a small effect may have a large benefit-cost ratio. The local government has access to a database containing the phone numbers of all voters. Norway has introduced electronic registration of votes in some municipalities. The voters still use the paper ballot, but the timing of all votes are registered electronically.[14]

[14] Importantly, the content of the vote is anonymous and not registered.

### Choice of identification strategy

In the introduction, I informally discussed how interference may pose a threat to the unbiasness of an RCT. In Section 2.5.4 I construct a model for the bias arising from interference, and I will show that this is a setting that is particularly vulnerable to such effects.[15] This is also a setting where it would have been politically and ethically controversial to randomize the set of receivers of the text message.[16]

### Timing and content of the message

Panagopoulos (2011) finds that encouragements to vote are more effective close to the election day. I randomly chose one of the days in the last week before the election, which turned out to be Friday, September 8, 2017. I also sent out the text messages in random time intervals within this day. 19 of the 38 electoral districts received the text message a few hours before the rest of the city. For technical reasons related to the transmission of the text message, the 38 districts were divided in two groups along geographical lines.[17] Two different versions of the text messages were sent out, and more details about the content of the text messages are provided in the Appendix.

[15] I show that, even if the true effect is zero at the population level, we will estimate a treatment effect larger than the expected effect in a small experiment when as few as 2.5% of voters are affected by Experimenter demand effects.

[16] See, e.g., *Bergens Tidende* 09.14.2017 and 09.19.2017 or *Dagsnytt 18* 09.19.2017 for discussions about whether this project can be perceived as interfering in the election. Importantly, there was no media coverage until after the election.

[17] The line was deliberately drawn to make the two areas as equal as possible on observable characteristics. The Appendix provides more details.

**Data**

Some part of the analysis requires comparisons with other cities, and in this case I will make comparisons with the 19 largest cities in Norway. This is the optimal size of set of comparison cities determined by using a simple optimization method that is described in the Appendix.[18] I also make comparisons across time. I use aggregate electoral data from the Norwegian Directorate of Elections and SSB for the past 6 elections.[19] Parts of the analysis requires daily (or hourly) voting data, and I only have access to these data for the 2017 election.

## 2.4 Results from the experiment

### 2.4.1 Descriptive results

I first compare the electoral turnout in Bergen with the turnout in the other Norwegian cities. The average turnout in the 19 largest cities was 77.3% in 2017, while the turnout in Bergen was 82.1%. This difference of 4.8 percentage points is unusually large. Using data from the past 6 elections (114 observations), we do not observe *any* city that shows a larger deviation from the average turnout than Bergen in 2017. The turnout in Bergen increased by one percentage point between 2013 and 2017, while the average turnout in the 19 cities dropped by 0.3 percentage points. A naive difference-in-difference estimate comparing Bergen and the national city-average is then given by 1.3 percentage points. It is hard to give these results a causal interpretation, as the parallel trends assumption is unlikely to be satisfied. The average city is not necessarily a suitable comparison for Bergen, but it is possible to

[18]But I will also show that we can change the set of comparison units without affecting the results.
[19]The data is available at ssb.no/statistikkbanken and valg.no.

synthesize a more suitable comparison unit.

### 2.4.2 Synthetic control method

The synthetic control method offers a standardized method for evaluating a policy intervention by constructing a synthesized version of the treated unit to serve as a counterfactual (Abadie et al., 2010). I construct a synthesized version of Bergen based on variables that predict electoral turnout. The lagged outcome variable (past electoral turnout) is generally a strong predictor of future turnout, and I also include the wage level (as a proxy for socio-economic factors) as well as the population of a city.[20] The synthesized version of the treated unit is then chosen to minimize some difference between the treated and synthesized unit prior to the treatment.[21]

I initially use the set of the $19$ largest cities Norway as the pool of donors and use data from the $6$ past elections. However, there is a better pre-treatment fit by changing the donor pool to the $23$ largest cities and the $5$ past elections.[22]

#### Bergen vs synthetic Bergen

Between the 2013 and 2017 elections, the electoral turnout decreased in the synthesized version of Bergen, while the actual turnout in Bergen increased.[23] The synthetic control estimate is given by 2.2 percentage points with 19 cities in the

[20] As a robustness test, I later change the set of variables and show that such a change does not affect the estimates.

[21] There are various differences to minimize, but I use a regression-based method that is used as default in the *Synth* package developed for Stata. This package can be found at https://web.stanford.edu/ jhain/software.html.

[22] The pool of 19 cities were chosen using an optimization algorithm, but it is also important to get a close pre-treatment fit. This raises a dilemma, as I want to the stay true to the pre-specified setup, but also want to get a close pre-treatment fit. I choose to report the results using both specifications.

[23] The weights for turnout in the synthesized version of Bergen are Trondheim (0.476), Tønsberg (0.253) and Oslo (0.274) with 19 cities in the donor pool. With 23 cities, the weights are given by Trondheim (0.338), Bærum (0.362) and Haugesund (0.3).

**(a)** Turnout with 19 cities in the donor pool.

**(b)** Turnout with 23 cities in the donor pool.

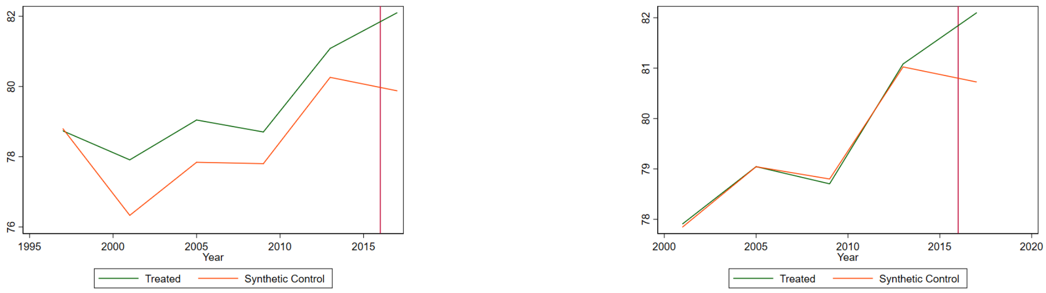



**Figure 2.1:** Turnout in Bergen versus synthesized Bergen.

donor pool, and 1.4 percentage points with the extended donor pool of 23 cities. Figure 2.1 compares the actual turnout in Bergen with the turnout in the synthetic version for both specifications of the model.

### Placebo tests across space and time

I follow Abadie et al. (2010) and use Placebo tests across time and space for inference. I first construct synthesized versions for the other cities, while removing Bergen from the pool of donors. Then I estimate the Placebo treatment effect for each city in the donor pool for all elections. I rank the effect for the treated unit according to this distribution. The effect in Bergen in 2017 is unusually large. Of the 19 cities in the past 6 elections (114 city-years), Bergen in 2017 provides the largest treatment effect. Using the extended pool of donors, the effect in Bergen is the second largest of 115 Placebo effects.

### Robustness

I here show that the results are robust to changes in the setup of the model. I change the set of variables and the number of elections used to construct the synthetic version of Bergen, and repeat the same analysis. The treatment effect for these versions of the model is given in Table 2.1. The Appendix provides graphical illustrations.

**Table 2.1:** Treatment effect for Bergen in 2017 for other versions of the model.

| Modification | Treatment effect | Ranking |
|---|---|---|
| 19 cities and six elections | | |
| Drop wage | 2.35 | 1 of 114 |
| Drop population | 2.46 | 1 of 114 |
| Drop one election | 1.88 | 3 of 95 |
| Drop two elections | 1.84 | 2 of 76 |
| 23 cities and five elections | | |
| Drop wage | 1.37 | 2 of 115 |
| Drop population | 1.37 | 2 of 115 |
| Drop one election | 1.60 | 2 of 92 |
| Drop two elections | 1.37 | 2 of 69 |

### 2.4.3 Treatment effects at time $t^\star$

The previous section argues that the electoral turnout was high in Bergen in 2017. Here I show that I can link the timing of the increase in turnout to the timing of the treatment. The text messages were sent out Friday, September 8, in the hours after 08:30 AM, and every voter had received a text message by 11:30 AM.

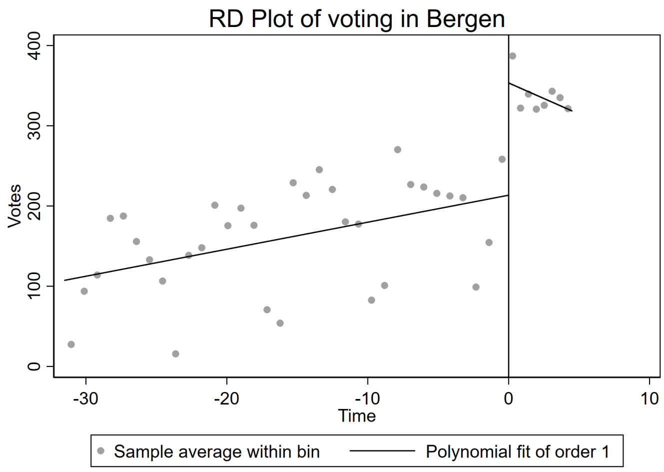
#### Voting in Bergen after the text message

The electoral turnout increased in Bergen after the treatment. Figure 2.2 shows a regression discontinuity (RD) plot of voting in Bergen before and after all voters received the text message.[24] The estimated jump at the threshold is given by 72 votes

[24] Using the rdrobust pacakage for Stata available from https://sites.google.com/site/rdpackages/rdrobust. RD allows for choices of bandwidth and local polynomials. I use the default options in the rdrobust package when I estimate the treatment effect. I only use observations from few days around the treatment, as Hausman and Rapson (2018) argue that using observations distant in time from the treatment may lead to biased effects.

per 15 minutes (p-value of 0.07), which indicates a treatment effect for the day of the treatment of around 1.3 percentage points. Further details about the measurement of the outcome variables and estimation of the size of the effect for the day of the treatment are given in the Appendix.

**Figure 2.2:** Votes cast in Bergen before and after all voters had received the text message. The vertical axis shows the number of voters in Bergen (measured per 15 minutes), while the horizontal axis shows time (measured in hours feasible for voting before and after the treatment).

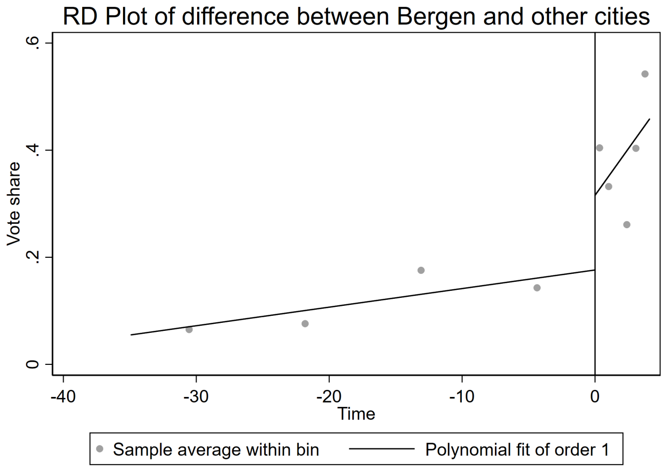



### Voting in Bergen after the text message relative to other cities

I adjust for potential time trends by transforming the outcome variable to the *difference* in turnout rates between Bergen and the other cities in Norway.[25] The election takes place at the same time in all cities, so underlying time trends may be relatively similar. I use turnout rates rather than the absolute number of votes, because the number of votes depends on the size of the city. The estimated treatment effect at the threshold is 0.27 percentage points per hour (p-value 0.04). This indicates a daily treatment effect of approximately 1 percentage point. Figure 2.3 shows a RD plot where difference between voting in Bergen and other cities is used as the outcome variable.

[25]The turnout rate is defined as the number of votes in a given time period divided by the number of voters that have not cast their votes.

**Figure 2.3:** The vertical axis shows the difference between turnout rates in Bergen and other cities before and after all voters had received the text message. The horizontal axis shows time (measured in hours before/after the treatment).



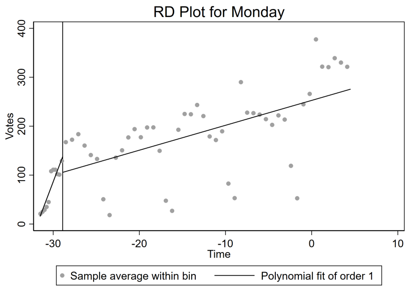

I also estimate the treatment effect using other methods than RD. There is a strong and positive correlation in Norwegian cities between turnout in the morning and later in the day for a given day. I use this relation and morning voting on the treatment day to predict turnout in Bergen for the afternoon of the treatment day in the absence of the text messages. This prediction leads to a treatment effect of 1.1 percentage points.

### Placebo tests

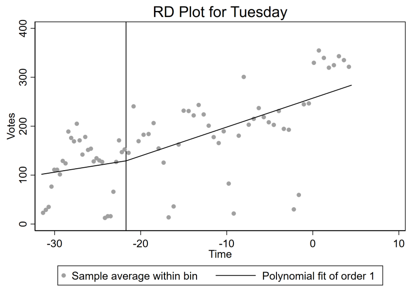
**Placebo tests using RD** All voters had received a text message by 11:30 AM, September 8. The key element behind Placebo tests is to check if we find treatment effects when there should not be any effect from the treatment. Figure 2.4 shows no signs of similar jumps at 11:30AM in the four days before the text messages were sent out.

I repeat this analysis when the outcome variable is the difference in turnout rates between Bergen and other cities. I do not find indications of jumps at other points in

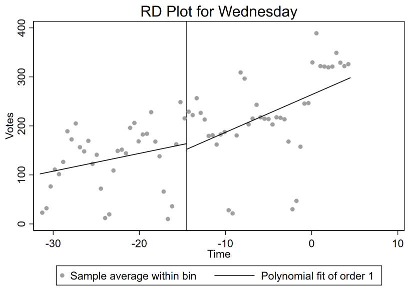
**(a)** Monday




**(b)** Tuesday




**(c)** Wednesday.




**(d)** Thursday

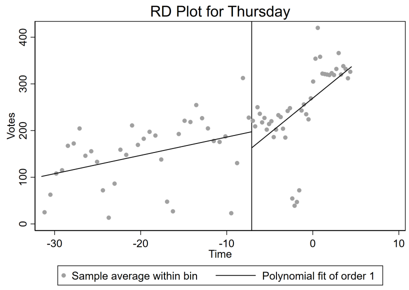



**Figure 2.4:** Placebo effects for the four days prior to the text messages.

time, which is shown in more details in the Appendix.

**Other Placebo tests** I also use the detailed dataset to compute further Placebo tests. In the time period after all voters received the text message, we observe a large increase in voting on the day of the text message compared to the day before. I split the time period between 11:30AM and 04:00PM in nine 30-minutes-intervals, and then I compare the increase in turnout after the text message with the daily change in turnout for the other $19$ days of early voting in Bergen. In *each* of these 9 time intervals, the increase in voting from day $t^{\star} - 1$ to day $t^{\star}$ is larger than for *all* of the other 18 daily changes. We do not observe this trend on the morning of the treatment day (before the text messages were sent out). I also use this method to analyze whether the change in Bergen after the treatment was large relative to observed changes in other cities. I find that the increase in turnout in Bergen between day $t^{\star} - 1$ and $t^{\star}$ is the second largest of $152$ daily changes in the $19$ cities. More

| | Votes | Turnout rate difference | Group 1 vs 2 |
|---|---|---|---|
| RD Estimate | 72.17* | 0.276** | 49.16*** |
| | (40.82) | (0.134) | (14.32) |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table 2.2:** The first column shows the RD estimate of number of votes in Bergen. The second column shows the RD estimate on the *difference* in turnout rates between Bergen and other cities. The third column shows the RD estimate on the difference in votes between the two parts of the city.

details and rankings of Placebo effects are provided in the Appendix.

### 2.4.4 Marginal variations in the timing of treatment

I use the variation of the timing of the text message to estimate further treatment effects. Half the districts of Bergen (group 1) received the text message between 08:30AM and 09:30AM. The other half (group 2) received the message between 09:30AM and 11:30AM. Figure 2.5 shows RD plots of the difference in votes cast between the two parts of the cities in the days before and in the hours after the first group received the text message. The left side of Figure 2.5 only includes voting data before the second part received the text message (before 09:30 AM), while the right side also includes later differences. We observe a large jump in the difference between the two parts of the city right after the text messages were sent out to the first group. The treatment effect is given by 49 votes per 30 minutes, and this effect is statistically significant (p-value: $0.001$). This treatment effect translates into an effect size of around $1.5$ percentage points for the treatment day. The treatment effects for the different models are summarized in Table 2.2.

**(a)** RD plot only including time periods before the second part received the text message.

**(b)** RD plot with more time periods.

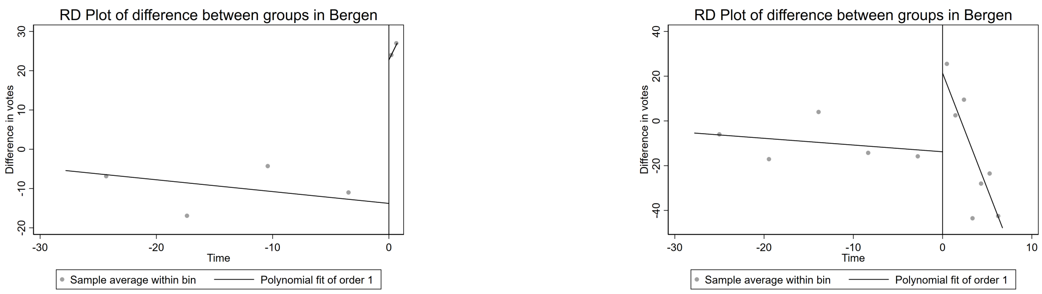



**Figure 2.5:** The vertical axis shows the difference in votes between the two groups (measured per 30 minutes), while the horizontal axis shows time (measured in hours before/after the treatment).

### Placebo tests

We do not observe such differences at previous points in time. I perform Placebo tests in time by changing the timing of the treatment to different hours in the preceding week. Then I estimate the RD effect for all these hours. The effect at the timing of the treatment is larger than all other effects.[26]

If I rather than using RD simply compare voting in the two parts of the city, there is also an unusually large difference after the treatment. There are $86$ and $62$ votes in group $1$ and group $2$ before 09:30AM on the day of the text message. This difference is larger than for all other days of early voting. The previous day there are respectively $27$ and $42$ votes at this time of the day. The difference-in-difference estimate is given by $(86 - 27) - (62 - 42) = 39$ in favor of group 1, and this is also larger than for all other days. Of the other 19 daily difference-in-difference estimates for Bergen, the second largest estimate in favor of group $1$ is $15$. The Appendix provides more details and further Placebo tests.

[26]The second-largest effect is 33 votes (with a p-value of $0.07$).

### 2.4.5 Underlying mechanisms

I also want to understand *why* a text message can increase the electoral turnout. The related literature discusses two main arguments for why such an encouragement can have an effect on the electoral turnout. The social connectedness theory argues that encouragements work when they are personal and social (Gerber and Rogers, 2009; Green and Gerber, 2015). This argument compares voting mobilizations to invitations to a party, and claims that the effect is stronger if others also attend the party. However, it has also been argued that voting mobilizations simply work as provision of information. The noticeable reminder theory (Dale and Strauss, 2009) argues that a reminder of the election can be sufficient to mobilize voters.

I used two different strategies to investigate the underlying mechanisms. I gave half of the voters a text message with a social content, while the others received a non-social content. To test the noticeable reminder theory, I simply asked people about their opinion of the informational aspects of text message, and elicited the effect of the text message by comparing surveys conducted before and after the treatment.

#### The effect of a text message with social content

19 of the 38 electoral districts received a text message using descriptive social norms to encourage voting.[27] The text message contained the (truthful) message that a majority of the people in Bergen vote, and that the turnout is higher in Bergen than in Trondheim and Oslo. The other half of the electoral areas received a non-social content. I do not have a set of comparison cities in this setting to evaluate whether the observed difference between the groups before and after the text message is large,

[27]These were the same districts that received the early text message.

so I construct hypothetical comparison units by using Monte Carlo simulation. This approach randomly allocates different electoral areas of the city to the different text messages, and then observes that approximately 66% of simulations lead to larger treatment effects than the true difference.[28] This result does not indicate that the two messages had different effects on turnout, and may suggest that the social aspects are less important for explaining why a text message may increase turnout. Importantly, I do not claim anything about the effect of social factors on voting. I just suggest that the effect of a text message on turnout may be through other channels.

#### Survey results about the informational content of the text message

A post-election survey was made by Respons Analyse on behalf of the local government of Bergen, using a representative sample of the population in Bergen ($n = 600$). Asking people whether the informational content of the text message affected their voting will arguably not lead to precise answers. Rather the survey asked people how they wanted to receive information about elections and other important public events. If the text message works as provision of information, we would expect people to want to receive such information through text messages. Multiple informational sources were listed as alternative sources of information (such as newspapers, webpages and billboards). 41% wanted to be informed through text messages, and this is more than for any other source of information. I cannot directly link this number to the text message. However, a similar survey was conducted some months prior to the election (using a different representative sample), and then only 4% wanted information through text messages. This may indicate that receiving this particular text message made the voters more positive to receiving such informational text messages.

[28]Details are provided in the Appendix.

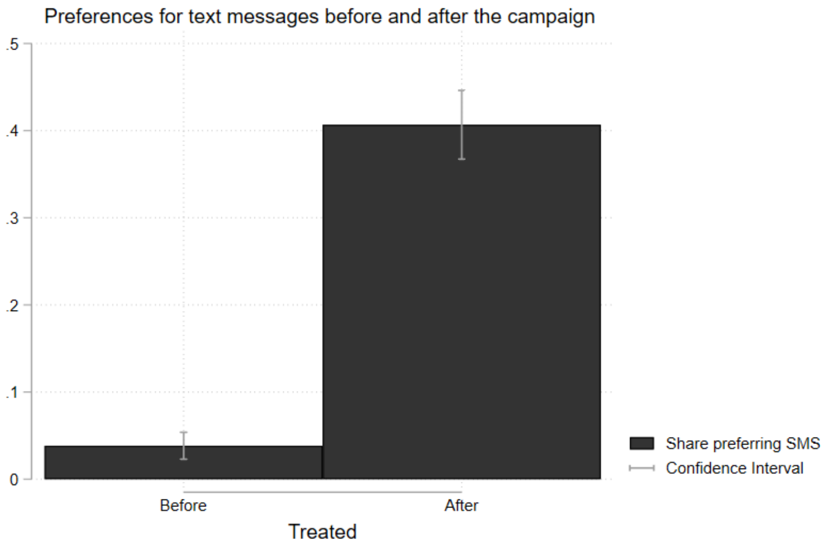



**Figure 2.6:** Share of voters preferring information through text messages before and after the text message

A necessary condition for a text message to act as a noticeable reminder is that voters notice the message. The survey asked voters, two months after the election, whether they remembered receiving a text message. 44 percent answered that they remembered the text message. This is a larger share than for other informational sources (such as newspaper advertising, billboards and information through social media). Generally, voters answer similarly across different demographic groups. The only exception seems to be that voters above 60 years are less interested in the text message as a source of information and are also less likely to have noticed the text message. This analysis is elaborated in the Appendix.

## 2.5 Causal inference with interference at the population level

In this section, I analyze how to draw causal inference at the population level when there is interference between units. Most of the experimental literature relies on the Stable Unit Treatment Value Assumption (SUTVA), which assumes there is no interference between units, such that RCTs will be unbiased at the population level (Cox, 1958; Rubin, 1980). In real life people may interact, and the spillover effect of a voting encouragement is one example of a violation of the no-interference assumption.[29]

The main insight from this section is to show that while interference generally is not a problem for small randomized experiments, the bias may be important when evaluating population-level effects. I also specify a model to analyze which parameter values that make the treatment effect particularly vulnerable to biases from interference.

### 2.5.1 Population-level treatment effect

The outcome variable of interest is given by $f_D[t,k]$, which is a function of time ($t$), treatment status ($D \in \{0,1\}$) and the number of recipients of the treatment ($k$). In this section, I suppose that only the number of recipients (and not the distribution) affects the other units. I also assume a homogeneous treatment effect, which means that the causal effect of this treatment at time $t+s$ when $k$ persons receive the treatment is given by

$$\alpha(t+s,k) = f_1[t+s,k] - f_0[t+s,0].$$

When SUTVA is satisfied, interference between units is ruled out, which means that an unbiased population-level treatment effect can be found using a smaller sample size. However, SUTVA is not a *necessary* condition for identifying unbiased

[29]See, e.g., Sacerdote (2001), Duflo and Saez (2003), Carrell et al. (2009), Bursztyn et al. (2014) and Dahl et al. (2014) for studies of peer effects. Rosenbaum (2007) provides an analysis of the assumption of no interference. Heckman (2005) points out other implicit assumptions behind SUTVA.

effects. Individual randomization with $k$ recipients estimates the treatment effect $f_1[t+s,k] - f_0[t+s,k]$, but usually the treatment group is small compared to the size of the population. Randomization also leads to an unbiased effect under assumption 2.

**Assumption 2.** $\lim_{\hat{k}\to k} f_D[t+s,\hat{k}] = f_D[t+s,k]\ \forall D \in \{0,1\}$.

If the treatment group is small ($\hat{k} \to 0$) compared to the size of the population, the probability of interference between a random person from the control group and the treatment group is small, which makes this assumption innocuous.

### 2.5.2 Bias from interference for the population-level effect

We may relax SUTVA and still identify unbiased treatment effects on a small scale as long as assumption 2 is satisfied. In this section, I want to investigate whether an unbiased population-level treatment effect can be identified using similar assumptions. Suppose we want to find the population-level treatment effect when assumption 2 is satisfied. This means that there are no spillover effects from the treatment to control if we let the number of treated units be small. For this to be the population-level treatment effect, we must *additionally* assume that the treatment effect is the same when there are few and very many recipients, and this is not satisfied if there are scale or equilibrium effects from the treatment. I want to find the effect when everyone receives the treatment, so an alternative is to let a large majority receive the treatment and compare with the outcome under control. When Assumption 2 is satisfied, it follows that $\lim_{\hat{k}\to N} f_1[t+s,\hat{k}] = f_1[t+s,N]$, which implies that we know the outcome under treatment. However, this leaves us with a small set of control units, and to get an unbiased treatment effect we need to assume that there are no spillover effects from treatment to control.

**Proposition 9.** *Assume that SUTVA is not satisfied. As long as assumption 2 is satisfied we can estimate an unbiased treatment effect if the treatment group is small, but we need to make more assumptions to estimate an unbiased population-level treatment effect.*

Proposition 9 is intuitive. In a small-scale experiment we want to compare $f_1[t+s, \hat{k}]$ and $f_0[t+s, 0]$. If $\hat{k} \to 0$ and the outcome variable is continuous, we will estimate something that approximates what we want to estimate. But for the population-level effect, we want to compare $f_1[t+s, N]$ and $f_0[t+s, 0]$, and there does not exist a value of $k$ that at the same time is close to 0 and $N$.

### 2.5.3 Magnitude of the bias

I have shown that in certain settings it may be difficult to estimate an unbiased population-level treatment effect when there is interference between units. But establishing the existence of a bias does not imply that the bias will have large effects on the estimates.[30] Here I suppose that the treatment is provided to a large majority of the population.[31]

In this section, we particularly want to understand the magnitude of the bias for the voting encouragements and similar type of encouragements and treatments. This implies that we think of the outcome as the share of units engaging in some action or behavior, which means that I restrict both $f_1[\cdot]$ and $f_0[\cdot]$ to take values between 0 and 1. I also assume that the true treatment effect, $\alpha$, is positive. The causal effect of the treatment when almost the entire population (of size $N$) is treated is given by $\alpha = f_1[t+s, N] - f_0[t+s, 0]$, but the effect we estimate is given

[30] On the other hand, estimating population-level treatment effects typically means that the sample will be large, and large sample sizes mean that the effect of even a small bias may be statistically significant.

[31] I set $\hat{k}$ so close to $N$ that I can assume $\lim_{\hat{k} \to N} f_1[t+s, \hat{k}] = f_1[t+s, N]$.

by $\alpha_N = \alpha + f_0[t+s, 0] - f_0[t+s, N]$, which means that the bias is given by $b(f_0) = f_0[t+s, 0] - f_0[t+s, N]$. A measure of the relative importance of the bias is then given by $y[b(f_0), \alpha] = \frac{|b(f_0)|}{\alpha + b(f_0)}$, which is the ratio between the absolute value of the bias and the observed treatment effect. I now analyze how $\alpha$ and $f_0$ affect the bias.

**Proposition 10.** *The relative importance of the bias ($y[b(f_0), \alpha]$) is decreasing in $\alpha$, and it will be increasing in $f_0$ if $b(f_0)$ is a monotonic and differentiable function that satisfies $b(0) = 0$.*

Proposition 10 has an intuitive explanation. A small expected true effect ($\alpha$) will make the bias larger relative to the true effect. For a given $\alpha$, a large $f_0$ increases the importance of the bias relative to the true effect. The problem for drawing inference when a large share of the population is treated, is that the control units are affected by the treatment, so a larger number of affected control units magnifies the bias. The importance of the bias is increasing in $f_0$ regardless of the direction of the bias.

### 2.5.4 The choice of identification strategy in the voting experiment

I use the insight from the previous section to analyze the consequences of using an RCT in the voting experiment.

**Small RCT** One possible strategy is to provide the treatment to a small share of voters. Assumption 2 then states that the outcome under control is unaffected by the treatment. However, we need to assume that the treatment effect is equal for a larger treatment group to argue that this is an unbiased population-level effect, and I have argued that the treatment effect is not necessarily the same at the population level.

**Population-level RCT** Another option is to increase the scale of the experiment and use a population-level RCT, by sending a text message to a large share of the voters. In this case, a bias may exist if the treatment affects the non-treated units. I use the model from Section 2.5.3 to evaluate the importance of this bias given the parameter values in a voting setting. The turnout in Norway is approximately given by $f_0 = 0.8$. The effect of a voting encouragement ($\alpha$) is typically found to be small, with a treatment effect between $0$ and $2$ percentage points. A small $\alpha$ and large $f_0$ means that the relative importance of the bias is large, which implies that this setting is particularly vulnerable to the effects of interference. We know from previous sections that there is clear evidence that voting encouragements spread within close relations. However, this is not the only bias that is magnified by these parameter values. Suppose a share $e$ of the voters in the control group are affected by Experimenter demand effects, which means that they choose not to vote if assigned to control by the government. If $e = 2.5\%$, we will estimate a treatment effect of $2$ percentage points when the true effect is zero.[32]

## 2.6 Conclusion

This paper has analyzed methods for drawing causal inference in settings where using an RCT for some reason is infeasible. This paper argues that we can find a causal effect by randomizing the *timing* and *order* of the treatment. I then employ such methods to find the effect of a text message from the government encouraging people to vote. A central finding of the paper is that it is possible to increase the electoral turnout by using a nudging text message. An advantage of randomization in time in this setting is the enhanced external validity. The experiment *is* an actual policy

[32] In this case, a population-level RCT leads to a treatment effect of $f_0 - (1-e)f_0 = ef_0$, which equals $0.02$ when $e = 0.025$ and $f_0 = 0.8$.

intervention, which means that the results may be more relevant for policymakers elsewhere.

# Appendix A. Details about RD measurements and estimation

### Measurement of the outcome variable

I measure the number of votes within certain time intervals, which means that I control the number of observations by adjusting the length of the time interval. We need observations from a certain number of periods after the treatment to be able to estimate a treatment effect. However, a central lesson from Hausman and Rapson (2018) is that RD in time may be biased when using observations distant in time from the cut-off.

When the outcome variable is the turnout rate difference between Bergen and other cities, we adjust for some of the underlying time effects, which means that I can use observations from a longer period of time. I split the votes in relatively large time intervals (60 minutes) and still have observations for enough post-treatment periods to estimate an effect.

When the outcome variable is the number of votes in Bergen, there may be biases by using votes cast a long time after the treatment.[33] Therefore, in this case, I rather choose to divide the time periods in shorter time periods (15 minutes) to have enough post-treatment periods to estimate an effect. I still have many observations within each 15-minutes-interval (in average around 200 votes for each 15-minutes period).

[33] In this case, "a long time" means a few days.

A disadvantage of using this outcome variable is that it may underestimate the true effect. Some voters had received the text messages earlier, which means that they may have voted in the 15 (or 30 or more) minutes preceding 11:30AM.

When the outcome variable is the difference between the two parts of the cities, I set the length of the time interval to 30 minutes. In this case, I adjust for time trends, but I also need detailed data, as there is only one hour where only group had received the text message.

### Estimating the treatment effect

The RD provides a point estimation. Any form of extrapolation in time relies on some further assumptions. To estimate the effect for the treatment day, I take two different approaches. In one approach, I simply extrapolate by assuming that the treatment effect is constant throughout the day. In the other approach, I estimate the treatment effect as the difference between the regression lines estimated by RD before and after the treatment.

**Votes cast in Bergen** The RD analysis using the number of votes in Bergen as the outcome variable leads to an estimated increase of 72 voters per 15 minutes. By aggregating this difference to the entire day, we get a treatment effect of approximately 1 percentage point.[34] Alternatively, I estimate the treatment effect by comparing the differences between the two regression lines from the RD plot in Figure 2.2. This leads to an estimated difference of approximately 1.4 percentage points.[35]

[34] The average voter is exposed for the treatment effect for 7.25 hours and Bergen has 199918 voters, which leads to an estimated treatment effect of $\frac{7.25*72*4}{199918} \approx 1.0\%$.

[35] The graph estimates an average difference of approximately 100 votes per 15 minutes for 7.25 hours, which indicates a treatment effect of 2900 voters, or $\frac{2900}{199918} \approx 1.45\%$.

**Difference in turnout rate between Bergen and other cities** The RD analysis using the difference between turnout rates in Bergen and other cities estimate a treatment effect of $0.27$ percentage points of voters per hour. This indicates a treatment effect of $1.4$ percentage points for the entire day given a constant treatment effect.[36] Comparing the two regression lines leads to an effect of approximately $1$ percentage point.

**Votes in group 1 versus group 2** The RD analysis estimates an increase of 49 votes per 30 minutes in group 1 relative to group 2, which indicates an hourly increase of 98 votes. For this analysis, I use observations *within* the one hour long time period where the first group received the text message as the post-treatment outcome. Because the text message to the first group was sent out uniformly through this hour, the voters in group 1 were, on average, exposed to the text message for half of this time period. The hourly estimated effect is then 196 votes for Group 1, which indicates a daily effect of $1.4$ percentage points.[37] Comparing the differences between the two regression lines leads to a treatment effect of approximately $1$ percentage point.

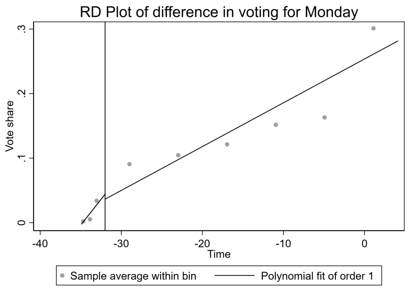
## RD Placebo tests for other days

Figure 2.7 shows RD plots for Placebo treatments taking place at the four preceding days. The outcome variable is the difference in turnout rates between Bergen and the other cities. This figure does not show any sign of discontinuities for the days prior to the text message.

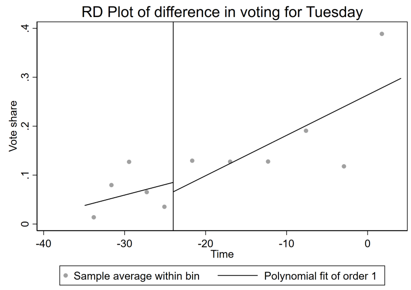
[36] There are 139,946 voters in Bergen that have not cast their votes at the time of the text message, which leads to a treatment effect of $\frac{0.0027*7.25*139946}{199918} \approx 1.4\%$.

[37] $\frac{98*2*7.25}{98457} \approx 1.4\%$.

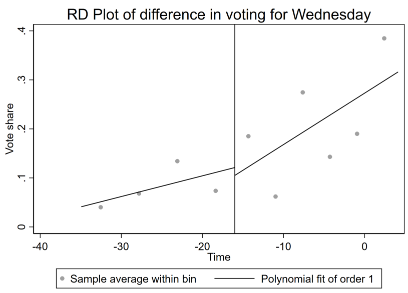
(**a**) Monday




(**b**) Tuesday




(**c**) Wednesday.




(**d**) Thursday

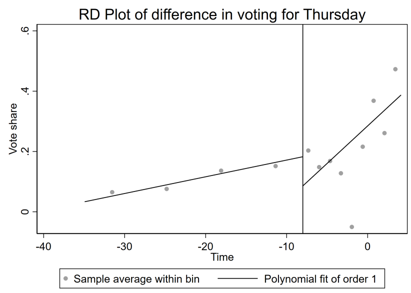
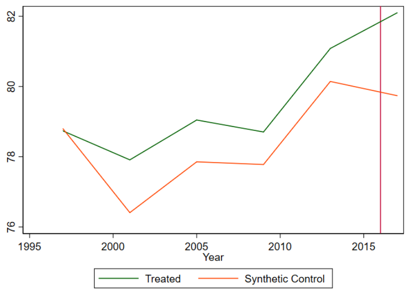



**Figure 2.7:** Placebo effects for the four days prior to the text messages.

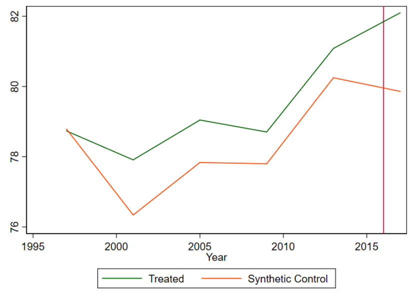
# Appendix B. Synthetic control estimates for different specifications

Figures 2.8 and 2.9 show the difference between voting in Bergen and the synthesized version of Bergen before and after the treatment, with respectively 19 and 23 cities in the pool of donors.

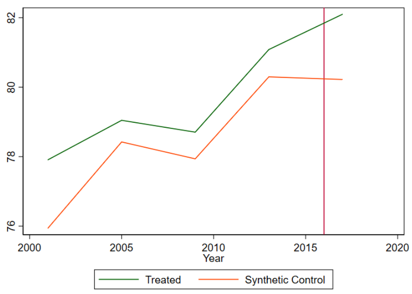
**(a)** Drop wage




**(b)** Drop population




**(c)** Drop 1997




**(d)** Drop 2001

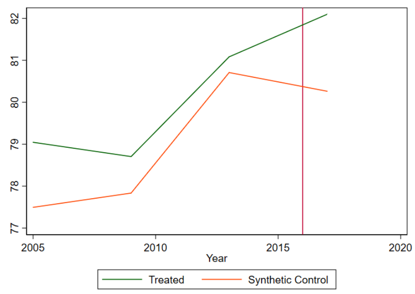
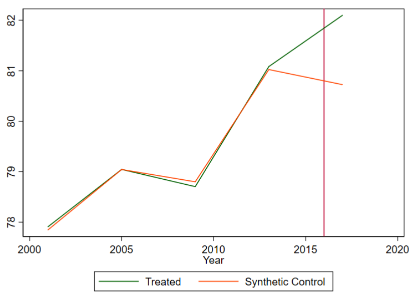
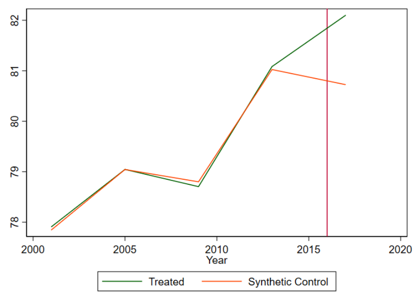



**Figure 2.8:** Synthetic control estimates for different specifications (19 cities).

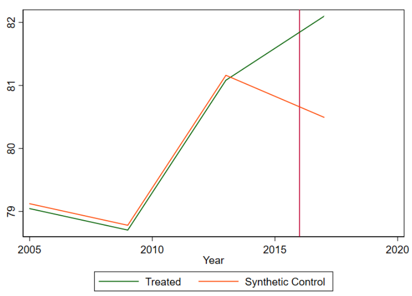
**(a)** Drop wage




**(b)** Drop population




**(c)** Drop 2001 election




**(d)** Drop 2005 election

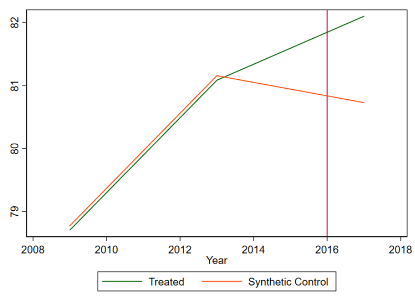



**Figure 2.9:** Synthetic control estimates for different specifications (23 cities).

# Appendix C. More details about the experimental design

## The set of comparison units

The size of the set of comparison cities was determined using the following rule. Suppose we rank cities in Norway ($j \in \{1, 2..\}$) according to the number of voters ($V^j$). The largest city of Oslo has $V^1$ voters and the second-largest city of Bergen has $V^2$ voters. A natural cut-off point ($M^\star$) is where the population difference between two adjacent cities is maximized, conditional on the set of cities being larger than a minimal set of cities. Formally, $M^\star$ is the solution to

$$M^\star = argmax \quad V^M - V^{M+1}$$

$$s.t \quad M \geq \underline{M}.$$

Here $\underline{M}$ is a minimal set of cities. If we want more than $\underline{M} = 8$ cities, then the difference is maximized at $M^\star = 19$.

## Division of the city

I divided the city of Bergen in two separated geographic areas (deliberately balancing on presumably important covariates such as previous turnout, education, population size and political affiliation) and randomly allocated the order as well as the content of the messages to the two parts. Bergen is divided in 38 electoral areas (*valgdistrikt*) that are primarily relevant for administrative purposes. 19 of these areas were selected as *Group 1* (the central and Western part of the city), while *Group 2* consists of the remaining 19 areas (the Northern, Eastern and Southern part of the city). The share of votes for the largest party in the previous election (*Høyre*) was used as a measure of political preferences. Table 2.3 shows that the two areas are relatively similar on observable covariates, but obviously, we do not know if they differ on non-observable characteristics.

**Table 2.3:** Covariate balance between the two parts of the city.

| | Group1 | Group 2 |
|---|---|---|
| Higher education | 37.6 | 36.7 |
| Turnout local election | 62.2 | 62.7 |
| Votes for Høyre | 32.5 | 34.0 |
| District population | 5520 | 5770 |

## Content of the message

The text message sent to Group 1 included the following message (translated from Norwegian).

1. Hi! Did you know that most people in your area usually vote? The turnout in Bergen was higher than in Oslo and Trondheim in the previous Parliamentary Election. Remember to vote on Monday!

   This message highlights descriptive social norms using the same approach as Gerber and Rogers (2009). The message sent to Group 2 included the following message.

2. Hi! Remember to vote on Monday. Your vote may determine the outcome of the election!

The Norwegian original version of the two messages are given below.

1. Hei! Vet du at de fleste i ditt område pleier å stemme? Bergen hadde høyere valgdeltagelse enn Oslo og Trondheim sist stortingsvalg. Husk å stemme på mandag!

2. Hei! Husk å stemme på mandag. Din stemme kan avgjøre valget!

# Appendix D. More Placebo tests from section 2.4

## Ranking of daily changes in Bergen in the week preceding the text message

Table 2.4 compares the increase between Thursday and Friday for each hour, given by $\tilde{\alpha}(t^{\star})$, with the changes between prior days.[38] Table 2.4 shows that the increase in the electoral turnout between $t^{\star}-1$ and $t^{\star}$ is larger than for all other days *for each* 30-minute period between 11:30AM and 04:00PM.

**Table 2.4:** Ranking of change in turnout among 19 daily differences

| Time | $\tilde{\alpha}(t^{\star}-3)$ | $\tilde{\alpha}(t^{\star}-2)$ | $\tilde{\alpha}(t^{\star}-1)$ | $\tilde{\alpha}(t^{\star})$ |
|---|---|---|---|---|
| Before $11:30AM$ | 5 | 3 | 2 | 10 |
| $12.00PM$ | 8 | 2 | 12 | 1 |
| $12.30PM$ | 4 | 10 | 16 | 1 |
| $1.00PM$ | 12 | 2 | 18 | 1 |
| $1.30PM$ | 19 | 4 | 15 | 1 |
| $2.00PM$ | 6 | 9 | 10 | 1 |
| $2.30PM$ | 2 | 19 | 6 | 1 |
| $3.00PM$ | 4 | 18 | 3 | 1 |
| $3.30PM$ | 6 | 4 | 7 | 1 |
| $4.00PM$ | 16 | 2 | 3 | 1 |

[38]For example, $\tilde{\alpha}(t^{\star}-1)$ is the change in voting between Wednesday and Thursday.

## Comparing voting before and after the text message across cities

Splitting the votes cast on a given day in two time intervals (before and after 11:30 AM), I compute the change in voting between two following days. Using the sample of 19 cities, two time intervals and the five days prior to the treatment leads to a set of $152$ Placebo effects.[39] The change in Bergen between the treatment day (Friday) and the day before is the second largest of these $152$ effects. Table 2.5 shows the ranking of the change in Bergen. In Table 2.5, I label the different effects according to the day and time of the day (morning/afternoon) to make it easier to read. We can for example observe that the change in Bergen between Wednesday afternoon and Thursday afternoon (labeled $\tilde{\alpha}(thu, aft)$) is relatively small. Table 2.5 indicates that the increase in voting Bergen between $t^{\star} - 1$ and $t^{\star}$ (labeled $\tilde{\alpha}(fri, aft)$) is unusually large compared to changes in other cities.

**Table 2.5:** Ranking of vote change for Bergen

| Time | Ranking |
|---|---|
| $\tilde{\alpha}(tue, mor)$ | 104 |
| $\tilde{\alpha}(tue, aft)$ | 48 |
| $\tilde{\alpha}(wed, mor)$ | 61 |
| $\tilde{\alpha}(wed, aft)$ | 30 |
| $\tilde{\alpha}(thu, mor)$ | 49 |
| $\tilde{\alpha}(thu, aft)$ | 50 |
| $\tilde{\alpha}(fri, mor)$ | 128 |
| $\tilde{\alpha}(fri, aft)$ | 2 |

## Marginal variation of timing

In the main model, I use 09:30 AM as the RD threshold. Here I will show the difference in votes for other points in time during the same day. Table 2.6 shows the hourly number of votes ($V_s^1$ and $V_s^2$) and the difference in turnout rates ($v_s^1 - v_s^2$) for

[39] I only use the week before to get enough observations for each unit; during the first weeks of early voting there are smaller cities that observe entire hours without voting.

the two parts of the city throughout the day of the treatment. Table 2.6 also shows the *ranking* of the treatment effect among 20 days of early voting in Bergen. There is no other day of early voting where the difference in voting in the morning is so large in favor of group 1. Around midday, the second group starts to mobilize many voters relative to group 1, and this difference is also unusually large. There are no other days where the afternoon difference in turnout between the two groups is as large as on the day of the treatment.

**Table 2.6:** Hourly voting distribution within Bergen

| Time | $V_s^1$ | $V_s^2$ | $v_s^1 - v_s^2$ | Ranking of $v_s^1 - v_s^2$ (of 20 days) |
|---|---|---|---|---|
| 10 | 159 | 108 | 0.0005 | 1 |
| 11 | 481 | 476 | 0.0001 | 10 |
| 12 | 683 | 664 | 0.0003 | 5 |
| 13 | 618 | 705 | -0.0007 | 18 |
| 14 | 612 | 668 | -0.0004 | 18 |
| 15 | 654 | 701 | -0.0003 | 19 |
| 16 | 605 | 690 | -0.0007 | 20 |

# Appendix E. Simulating the effect of different messages

Here I provide more details about how I use simulation to investigate whether the difference in voting between the parts of the city is large. There are $K$ electoral areas and areas $k \in \{1, 2...\hat{k}\}$ belong to group 1. I compute the share of votes in Bergen cast in one of these areas before ($t = b$) and after the treatment ($t = a$). The difference, given by $\nu_{true}$, is positive if group 1 mobilizes relatively more voters after the treatment.

$$\nu_{true} = \frac{\sum_a \sum_{k=1}^{\hat{k}} V_{tk}}{\sum_a \sum_{k=1}^{K} V_{tk}} - \frac{\sum_b \sum_{k=1}^{\hat{k}} V_{tk}}{\sum_b \sum_{k=1}^{K} V_{tk}}$$

The true effect is given by

$$\nu_{true} = \frac{48332}{98726} - \frac{31873}{64446} = -0.005.$$

Then I draw a random set of areas $\tilde{k}$ as a simulated group 1 and $K - \tilde{k}$ as group 2. I want the two groups to be balanced in terms of covariates ($x_i$). I only keep the simulations that ensure balance in education and previous turnout, by only using the simulations where the covariate difference between the two groups is smaller than a certain requirement.

I simulate 100,000 allocations and compare the estimates with the true effect. The average simulated effect (in absolute value) is twice as large as the real effect. 66% of simulated differences are larger than the true effect if we demand almost complete balance (only 2 % of the simulation will satisfy the requirement). A relaxation of the strong balance restriction increases the share of treatment effects larger than the true effect to 67%.

# Appendix F. Who are encouraged by the text message?

I analyze how different demographic groups answer the survey. Voters above 60 years tend to be less interested in text messages as a source of information about the election, while the interest in personal mail as an information source is similar across the age groups, as shown in Figure 2.10. While 45% of respondents below the age of 60 wanted information through text messages, only 25% above 60 years preferred this source of information.

While 50% of the respondents younger than 60 years remembered the text message, the share of respondents above 60 years remembering the text message is given by 27%. More voters above 60 years remembered information about the election in newspaper advertisements, as shown in Figure 2.11. This may indicate that the effect of a text message on turnout is strongest among young voters - although I cannot causally establish such an age effect. Otherwise, the answers are relatively similar across most segments of voters. Women and men have similar preferences about the message, and so do people with high and low income and education. There does not seem to be important geographical variations.

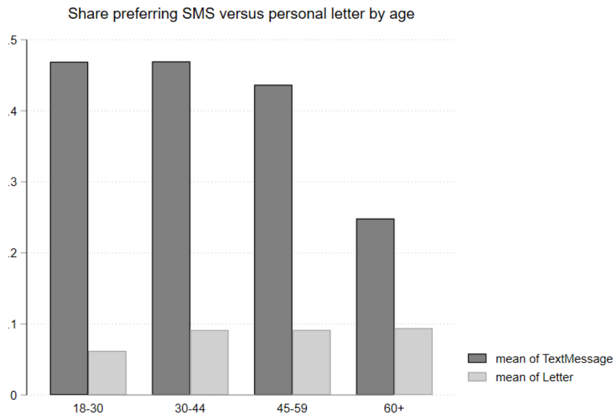



**Figure 2.10:** Share of voters preferring information through a text message and personal communication (through a letter) for different age groups.

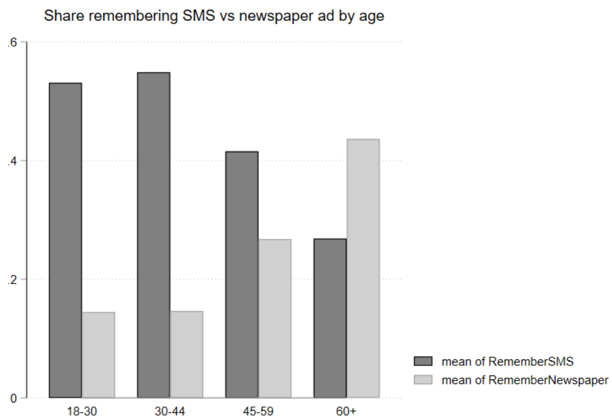



**Figure 2.11:** Share of voters remembering the text message and electoral advertisements in newspapers for different age groups.

# Chapter 3

# Under-reporting of Crime

Ole-Andreas Elvik Næss[1]

**Abstract**

Under-reporting of crime is widely regarded as a large problem. I construct a model to explain the observed patterns of crime reporting, and I find that crime reporting is lower than optimal. The victims do not internalize the effect of an additional report on the payoff of other victims. I show that introducing mechanisms that incentivize crime reporting will increase social welfare in certain settings. Remarkably, I also find that it may be optimal to make the reporting process costly and uncertain to discourage false reporting and to sustain an equilibrium where a report is an informative signal of guilt (*JEL*: D82, H80, K14).

## 3.1 Introduction

Less than one fourth of all sexual assaults in the US are reported to the police (Luce et al., 2010), and a low level of crime reporting is regarded as a large problem

[1]I would like to thank seminar participants and discussants in Bergen for helpful comments and discussions. I would like to extend a special thanks to Tore Ellingsen and Eirik Gaard Kristiansen for tremendous guidance and support. The usual disclaimer applies.

for societies.[2] Recent high-profile cases provide examples of the extent of under-reporting. The famous English media personality Jimmy Savile and the American film producer Harvey Weinstein committed sexual abuse over a period of several decades. While the empirical evidence generally presents a dismal picture of the extent of under-reporting of crime (Fisher et al., 2003; Mengeling et al., 2014), the above examples also show that certain circumstances lead to sharp increases in reporting. Starting from October 2017, more than 100 women accused Harvey Weinstein of sexual harassment.[3] These decisions are probably not independent, as it is statistically unlikely that 100 victims independently decided to report after remaining silent for years. In the months following the Savile case, England observed a large increase in accusations made against other profiled offenders, as well as an increase in reporting of crime in the general population.[4] After the Weinstein case there was a similar development, with the #MeToo movement as a result.[5]

This paper has two main contributions. The first contribution is to provide an explanation for why there is a low level of crime reporting, but also a high level of volatility in reporting rates. The second contribution is to analyze if it is possible to use the insights from victims' reporting decisions to construct mechanisms to increase social welfare.

I first build a model to understand crime reporting. *Victims of crime* decide whether to report a crime, and there may exist agents making *false reports*. The victims want to convict their offenders, but the victims face heterogeneous costs of taking the case to court. A *Bayesian jury* collects the reports, and convicts a defendant if there is

[2]Besides the costs of criminals not being held responsible, more reporting may also deter crime (Goldberg and Nold, 1980; Green et al., 2019; Iyer et al., 2012). A small increase in reporting may have large benefits given that Peterson et al. (2017) estimate the lifetime costs of rape victims in the US to 3 trillion dollars.

[3]https://edition.cnn.com/2020/01/23/us/witnesses-harvey-weinstein-trial/index.html.

[4]https://www.bbc.com/news/uk-28340196.

[5]Levy and Mattson (2019) find that the #MeToo movement led to an increase in crime reporting of 14%, and also find that the effect is persistent. Rotenberg and Cotter (2018) find that the #MeToo movement also increased crime reporting in Canada.

sufficient evidence that the defendant is guilty. The set-up of the model primarily applies to settings where hard evidence is unavailable and at least some criminals are guilty of more than one crime.[6]

I show that a Bayesian jury convicts a defendant after observing more than a certain number of reports. The decision rule of the Bayesian jury creates an issue of information asymmetry for the victims, because the victims do not know if there are other victims of the same criminal. There are spillover effects associated with crime reporting, because an increase in reporting by other victims will make it more likely to make a conviction. A more subtle spillover effect to independent cases arises because the reporting strategies of other victims also affect the number of reports necessary for a Bayesian jury to make a conviction.[7]

I show that a lower than socially optimal level of crime reporting will occur in equilibrium. The reason is that victims fail to internalize the positive externalities associated with crime reporting. An increase in crime reporting benefits other victims, which means that under-reporting of crime in equilibrium may be interpreted as under-provision of a public good.

I let a social planner control institutional features of the judicial process, such as the necessary burden of proof for conviction, monetary payments and the set of reports taken to court. An implication of under-reporting of crime is that measures aiming to increase crime reporting may unambiguously increase social welfare, while other measures typically involve trade-offs. Lowering the burden of proof will for example increase the probability of convicting guilty defendants, but will also increase the

[6]The model can also be applied in settings where victims have access to hard evidence and all criminals are guilty of exactly one crime, but then the spillover effects identified in this paper will not affect the reporting decisions.

[7]A Bayesian jury weighs the probability of observing a given number of reports against a guilty and innocent defendant. When it is common knowledge that the level of truthful reporting is low, observing one report may not be a strong signal of guilt. However, when the jury knows that more true victims report, the probability of observing one report against a guilty defendant may increase.

probability of making wrongful convictions. However, although a small increase in reporting increases social welfare, it does not necessarily imply that a social planner is able to implement mechanisms that *only* incentivize more truthful crime reporting.

When there are few or exogenous false victims, mechanisms that incentivize reporting by reducing the costs and uncertainty associated with the reporting process will increase social welfare. I show that social welfare is maximized using a mechanism where victims deliver reports to a principal, and then the principal takes the case to court if and only if there is sufficient evidence for conviction. In the law literature, such a mechanism has been labeled an Information Escrow by Ayres and Unkovic (2012). This mechanism eliminates the information asymmetry by letting a report be pivotal only in circumstances where a victim wants to report. *The Callisto Project* employs the intuition underlying the Information Escrow, by designing an online crime reporting mechanism where the authorities are alerted if and only if another victim also files a report.[8]

Surprisingly, when taking into account the possibility of false reporting, the Information Escrow is not only a sub-optimal mechanism, but may also substantially decrease social welfare relative to the equilibrium outcome. The Information Escrow attracts more false reporting, which means that a Bayesian jury knows that a given number of reports becomes a weaker signal of guilt. When there is much false reporting, the outcome using the Information Escrow may be given by an uninformative cheap talk equilibrium, where the Bayesian jury makes close to zero convictions.

Although the information asymmetry and the costs of taking the case to court are the sources of under-reporting in this model, I show, somewhat paradoxically, that both of these features may be welfare-improving in the presence of false reporting. The reason is that an issue of information asymmetry also occurs between true and false

[8] https://www.projectcallisto.org/

crime victims. Unlike false victims, the true victims know that they report against a criminal agent, which means that the probability of observing multiple reports is larger for true victims.

I then analyze how this information asymmetry can be used to improve social welfare. I construct a *pledging mechanism*, where all victims pay a fee, or a pledge, for reporting a crime, and the fee is returned (such that the budget-balance is satisfied) if the defendant is convicted. To sustain a separating equilibrium without false reporting, this payment must be combined with uncertainty about the outcome in court. The reason is that in a separating equilibrium, the posterior probability of guilt equals one after observing one report. Using the pledging mechanism, the jury commits to convict with a probability lower than one after observing one report. Because the probability of observing more than one report is higher for true victims, this mechanism may create a separating equilibrium. If the size of the pledge can be arbitrarily large, I show that this pledging mechanism maximizes social welfare.

Remarkably, the arguments underlying the pledging mechanism rationalize several apparently sub-optimal features of the design of systems for crime reporting. The mechanism provides an argument for why reporting a case needs to be a costly and uncertain process. First, the mechanism argues that it needs to be costly to take a case to court to reduce false reporting. A second and more subtle argument is that uncertainty in the conviction decision is *necessary* to implement a separating equilibrium in this model. The mechanism also rationalizes why it is difficult for a victim to convict a defendant after one report ("one person's word against another's"), although there often is a low level of false crime reporting.[9] This relation may seem puzzling, as one report is a strong signal of guilt when there are few false reports. However, this model rationalizes such behavior. In this model, the reason for acquittal is not necessarily that the jury does not believe the victim, but acquittal will also

[9]Lisak et al. (2010) and Spohn et al. (2014) find false reporting rates between 2% and 10%.

occur because a low probability of conviction after one report may be necessary to sustain a separating equilibrium without false reporting.

**Related literature** This paper is related to a recent paper by Lee and Suen (2019), who analyze the credibility of early versus late crime reports given unverifiable information. A working paper by Cheng and Hsiaw (2019) analyzes under-reporting of sexual misconduct against a manager. I take a different approach by focusing on the equilibrium outcome given a Bayesian jury and welfare effects of different mechanisms for crime reporting. Pei and Strulovici (2019) also study credibility of reports, but their main focus is on the incentives of criminals. More generally, this paper builds on different branches of the economic literature. A large literature, starting with the seminal contributions from Becker (1968) and Ehrlich (1973), analyzes the incentives for criminals to engage in crime.[10] Positive externalities from actions to prevent crime have been shown empirically (Ayres and Levitt, 1998; Cook and MacDonald, 2011). This paper also builds on the coordination games literature (Schelling, 1960) and the works on voluntary provision of public goods. Ayres and Unkovic (2012) argue that conditioning the outcome on the reporting of others may increase the incentives for reporting.[11] This paper analyzes settings with unverifiable reports, which implies that this paper is related to Chassang and i Miquel (2018), and the model in this paper also relates to the global games literature (Carlsson and Van Damme, 1993; Morris and Shin, 1998).

The building blocks of the model in this paper are empirically supported. Interviews with crime victims indicate that an important reason for reporting is to convict the offender (Patterson and Campbell, 2010). A Canadian survey showed that 46% of

[10] This literature is reviewed in Paternoster (2010), Nagin (2013), Draca and Machin (2015) and Chalfin and McCrary (2017). See also Balkin and McDonald (1981), Cook (1986), Furlong (1987), Ehrlich (1996) and Allen (2007).

[11] See e.g., Palfrey and Rosenthal (1984), Bagnoli and Lipman (1989) and Tabarrok (1998) for other settings where conditioning the outcome on the actions of others may be beneficial.

sexual assault victims list a lack of evidence as a reason for not reporting, while 34% did not want the hassle of dealing with the court process.[12] There is also clear evidence that there is much variation in the distribution of crime. Using a sample of 120 rapists Lisak and Miller (2002) find that 37 % are guilty of only one crime, while the average number of victims for multiple offenders is 5.8. Ayres and Unkovic (2012) argue that it may be difficult for a harassment report to prevail in the absence of other reports.[13]

This paper is structured as follows. The set-up of the model is described in Section 3.2. I analyze a basic model where only true crime victims are endogenous in Section 3.3. Section 3.4 provides an analysis of the general model, while Section 3.5 presents a numerical example of the model.

## 3.2 The model

In this section, I build a model of crime reporting. The population consists of a continuum of agents of size one, where some agents engage in crime. Victims may report the crime, and there may also exist agents making false crime reports. The police collect the reports and take the case to court. In court, a Bayesian jury (or a judge) chooses whether to make a conviction.

**Victims** The set of actions for a victim $i$ is whether to report the crime to the police, and this reporting decision is labeled $r_i \in \{0, 1\}$.[14] A victim faces costs $c_i$ if the case is taken to court. Such costs may, for example, be related to the costs of taking part

[12] Statistics Canada "Self-reported sexual assault in Canada, 2014", available from https://www150.statcan.gc.ca/n1/pub/85-002-x/2017001/article/14842-eng.htm

[13] Anecdotal evidence suggests that more reports increases the probability of conviction. A 2020 article from The Economist analyzes this relation: https://www.economist.com/international/2020/01/04/why-so-few-rapists-are-convicted.

[14] An indifferent agent is assumed to choose $r_i = 1$.

in the court process or a fear of retaliation from the offender.[15] A victim gets benefits $b$ if she reports and the offender is convicted.[16] In some settings, $b$ can be interpreted as monetary benefits, but $b$ may also be non-pecuniary benefits from convicting a guilty offender. Potential victims get a payoff of $V$ (normalized to $0$) by *not* being a victim of crime. The utility of a realized victim of crime is then given by

$$u(c_i, r_i) = \begin{cases} b - c_i & \text{if } r_i = 1, Conv, \\ -c_i & \text{if } r_i = 1, NotConv, \\ 0 & \text{if } r_i = 0. \end{cases}$$

The above expression displays the payoff, $u(c_i, r_i)$, as a function of the victim's costs of taking the case to court, the reporting strategy and the decision in court. There is arguably also a sunk cost of being a victim of crime, but introducing such a cost will not affect the analysis because the probability of being a victim of crime is constant in this model.

Each victim knows his or her own costs of taking the case to court, but does not know the reporting costs of other victims. It is common knowledge that the cost is larger than zero and distributed according to some continuous distribution $c_i \sim F(\cdot)$, where the minimum and maximum costs are given by $\underline{c}$ and $\overline{c}$. I assume that $\overline{c} < b$, which means that all victims want to report a crime if they are certain that a report leads to conviction.

Additionally, each agent faces a possibility of being wrongfully convicted of a crime,

[15] In practice there may also be psychological costs from reporting to the police. Introducing such costs changes some of the notation, but a small cost of reporting to the police will generally have minor consequences for the predictions from the model. The retaliation costs may also be modeled to be decreasing in the number of reports, which would magnify the spillover effects.

[16] Another option is to include a positive payoff also when the offender is convicted and the victim does not report. But if this payoff is large, we would observe few reports against criminals that are likely to get convicted.

which leads to a cost $C_w$.

**False victims** A share $\gamma$ of the population potentially engage in false reporting of crime. These agents receive a payoff of $b$ from convicting one randomly drawn agent from the population.[17]

**Criminals** A share $\pi$ of the population are criminals, and these agents engage in some given distribution of crime.[18] The probability that each agent is a victim of crime is labeled $\Pi$.

**Bayesian jury** A Bayesian jury collects the reports from the victims and chooses whether to make a conviction. The defendant is considered innocent until proven guilty, which can be translated into saying that the defendant is acquitted if the posterior probability that he is guilty conditional on the collected reports is less than $\overline{p}$. The Bayesian jury only observes unverifiable reports.[19]

**Timing of the game**

0. Nature draws the type $c_i$ for each $i$. A share $\gamma$ of agents are drawn to get a payoff from convicting a randomly drawn agent, while a share $\pi$ engage in crime.

[17]Modeling false reporting this way is consistent with at least two different motivations for false reporting. Agents may falsely report a crime if they want publicity, or if they hold grudges against others, and these grudges are uncorrelated with other features. For technical reasons, I assume that at least a share $\epsilon$ of agents engage in false reporting.

[18]I do not make any assumptions about the distribution if crime, but the model will not contain spillover effects for the particular, degenerate distribution where all criminals engage in exactly one crime.

[19]It is straightforward to incorporate hard evidence (e.g. DNA) into the model. Given that $b > c_i \forall c_i$, all victims with access to hard evidence choose to report and the spillover effects analyzed in this paper will not be important for these victims.

1. Each true and false victim decides whether to report a crime.
2. The police take the case to court.
3. The jury convicts a defendant if the posterior probability of guilt is larger than $\overline{p}$.

The equilibrium concept employed will be Perfect Bayesian Equilibrium. Section 3.3 analyzes a basic version of the model where only the behavior of true victims is analyzed, and then the equilibrium concept will be Bayesian Nash Equilibrium.

### 3.2.1 Social welfare

A social planner generally wants to maximize some weighted average of the welfare of all agents in the population. This raises the question of whether the payoff arising from engaging in false crime reporting should be included in the social welfare. I choose *not* to include the payoff arising from such activities in the social welfare function. Otherwise, the optimal outcome may be to encourage false reporting if the payoff from engaging in false crime reporting is sufficiently large, and this does not match observed policy objectives. These agents may also be true victims of crime, and in this case the payoff is included in the objective function of the social planner. Each agent is a victim of crime with probability $\Pi$, which leads to a payoff of $u(c_i, r_i)$. Non-victims receive a utility of $V = 0$, while the cost of being wrongfully convicted is $C_w$. All agents are equally likely to be wrongfully convicted of a crime, and this probability is labeled $s_w()$.[20] For each agent, the payoff is given by $\Pi u(c_i, r_i) - s_w()C_w$. The social welfare is found by aggregating the payoff for all agents, which leads to

[20] This probability is a function of the strategies made by the false victims, as well as the burden of proof necessary for conviction.

$$W = \int_{\underline{c}}^{\bar{c}} [\Pi u(c_i, r_i) - s_w()C_w]F'(c_i)dc_i. \tag{3.1}$$

### 3.2.2 The choice of a social planner

The social planner determines the institutional features of the reporting and conviction process, but cannot influence the agents' types. The social planner can control which cases that are taken to court, and which decisions that are made in court. I allow the social planner to impose fees or payments for reporting.[21] Given that the social planner does not observe the types of the different agents, all mechanisms must be incentive-compatible.

## 3.3 Analysis of a basic model

I first analyze a basic model, where I only focus on the behavior of true victims of crime. A numerical example of this model is provided in Section 3.5. I here assume a constant share $\gamma$ of agents providing false crime reports. In this section, the conviction decision of the jury is also simplified by assuming that there will be a conviction after observing $\tilde{k}$ or more reports. The probability for a victim choosing to report that there will be made $\tilde{k}$ or more reports is labeled $p^v(q) = P[k \geq \tilde{k}|q]$ when a share $q$ of victims report a crime.[22]

**Proposition 11.** *Let $q^\star$ be the solution to $q^\star = F[bp^v(q^\star)]$, where $F(\tilde{c}) = q^\star$. A Bayesian Nash Equilibrium is given by*

[21] Such payments are assumed to be financed by lump-sum taxation.

[22] $p^v(q)$ is increasing in $q$, because more victims choosing to report will increase the probability of observing more than $\tilde{k}$ reports.

$$r(c_i) = \begin{cases} 1 & \textit{if}\, c_i \leq \tilde{c}, \\ 0 & \textit{if}\, c_i > \tilde{c}. \end{cases}$$

Proposition 11 says that only victims with low costs choose to report. The threshold for reporting increases in the share of reporting victims, as more reporting increases the probability of conviction for other victims.

### Comparative statics

I here analyze how changes in the parameter $b$ affects the equilibrium. The direct effect is given by $F'(\cdot)p^v(q^\star)$, but an increase in $b$ affects all victims, which means that the equilibrium effects must be taken into account. Implicitly differentiating $q^\star(b) = F[bp^v(q^\star(b))]$ leads to

$$\frac{\partial q^\star}{\partial b} = \frac{1}{1 - bF'(\cdot)\frac{\partial p^v(q^\star)}{\partial q^\star}} F'(\cdot)p^v(q^\star).$$

The direct effect is multiplied with a factor given by $\frac{1}{1-bF'(\cdot)\frac{\partial p^v(q^\star)}{\partial q}}$. This equilibrium effect occurs because a larger $b$ also increases reporting by other victims, which will increase the conviction probability. The size of this multiplier depends on how much the conviction decision is affected by the reporting strategies of other victims ($\frac{\partial p^v(q^\star)}{\partial q}$). This reasoning can cast some light on the observed spillover effects in crime reporting. Small changes in parameter values may have large effects in equilibrium when the payoff depends on the strategies of other victims.

### Social welfare in equilibrium

In this section, I analyze the effect of a small increase in reporting in equilibrium.

**Proposition 12.** *Suppose there is an equilibrium where $q^\star > 0$. Consider a reporting strategy given by*

$$r^\star(c_i) = \begin{cases} 1 & \text{if } c_i \leq \tilde{c} + \Delta, \\ 0 & \text{if } c_i > \tilde{c} + \Delta. \end{cases}$$

*For any small $\Delta$, $W_{r^\star(c_i)} \geq W_{r(c_i)}$.*

Proposition 12 shows that a small increase in reporting will increase social welfare, which implies that there is under-reporting of crime in equilibrium. In an interior equilibrium, there will exist a victim that is close to indifferent between reporting and not reporting, and marginally chooses not to report. If this victim rather chooses to report, the probability of conviction increases for other victims. This implies that there is a positive externality from reporting a crime, where a small increase in reporting increases social welfare by

$$\frac{\partial W_{r(c_i)}}{\partial q} = \Pi[b\frac{\partial p^v(q^\star)}{\partial q}q^\star]. \tag{3.2}$$

There is also another, related issue of information asymmetry in this model. The victims choose whether to report without knowing the number of other victims. One victim may choose to report, although there is not enough evidence for conviction. Another victim does not report, although there would have been enough evidence for conviction. Social welfare would increase if these two victims switched strategies.

### 3.3.1 Mechanisms to increase social welfare in the basic model

I now analyze how different measures can be used to incentivize more crime reporting.

#### Adjusting the burden of proof

The social planner can adjust $\tilde{k}$, which is the necessary number of reports for conviction. For victims of crime, the utility is decreasing in $\tilde{k}$, and $\tilde{k} = 1$ is optimal. When $\tilde{k} = 1$ all reports lead to convictions, which means that the issue of information asymmetry disappears. However, $\tilde{k} = 1$ also leads to more wrongful convictions. The effect of $\tilde{k}$ on the number of wrongful convictions is given by Lemma 6.

**Lemma 6.** *The probability of being wrongfully convicted is* $s_w(\tilde{k}) = 1 - \sum_{i=0}^{i=\tilde{k}-1} e^{-\gamma} \frac{\gamma^i}{i!}$, *and* $s_w(\tilde{k})$ *is decreasing in* $\tilde{k}$.

Lemma 6 says that the probability of observing $\tilde{k}$ or more reports follows a Poisson distribution.[23] Lowering $\tilde{k}$ increases the likelihood of observing at least $\tilde{k}$ reports. Hence, the optimal choice of $\tilde{k}$ involves a trade-off between these effects. Whether or not adjusting the burden of proof increases social welfare depends on the welfare weight on avoiding wrongful convictions ($C_w$) relative to convicting guilty offenders.

#### Payments

A consequence of Proposition 12 is that a social planner wants to implement a marginal increase in crime reporting. A mechanism involving a monetary payment of

[23] In a population with $\gamma n$ agents giving false reports, the probability of observing more than $\tilde{k}$ reports follows a binomial distribution with probability $\frac{1}{n-1}$. The Poisson limit theorem says that such a distribution can be approximated by a Poisson distribution as $n \to \infty$. In this model, the population consists of a continuum of agents, which means that I will use such approximations.

$\epsilon$ will for example unambiguously increase social welfare. Although the payment is not necessary to induce reporting from victims with low costs of reporting, incentive compatibility forces the planner to extend the offer to all reporting victims. Paying victims to report internalizes the positive externality from a crime report. However, this mechanism will not solve the issue of information asymmetry. Some victims may choose to report although there is insufficient evidence for conviction.

### Optimal mechanism

I now show that there will be a particular mechanism that maximizes social welfare for any given value of $\tilde{k}$.

**Definition 1.** *An Information Escrow consists of victims choosing whether to report a crime to a principal. The principal takes the case to court if and only if there is sufficient evidence for conviction.*

**Proposition 13.** *The Information Escrow maximizes social welfare for any $\tilde{k}$.*

Regardless of the welfare weight of convicting guilty offenders relative to avoiding wrongful convictions, this mechanism maximizes social welfare for given values of $\tilde{k}$. The intuition for why this mechanism is optimal is straightforward. The interests of the victims are aligned. *All* victims want to take a case to court conditional on conviction, while no victims want to take a losing case to court. The victims are unable to coordinate with other potential victims of the same offender, but the principal is able to solve the coordination problem.

This mechanism unambiguously increases social welfare in the basic model, but still we do not observe a widespread use of such mechanisms in real-life judicial systems. A potential reason may be that the basic version of the model does not include the effect of the mechanism on other agents in the judicial process.

## 3.4 Analysis of the general model

Section 3.3 analyzes the behavior of true victims of crime, but here I extend the analysis by making the behavior of the jury and the false victims endogenous. I solve the model by backward induction. First, I analyze the conviction decision of a Bayesian jury, and then I analyze optimal reporting behavior by true and false victims conditional on this conviction decision. Section 3.5 provides a numerical example.

### 3.4.1 Analysis of a Bayesian jury

I first analyze the conviction decision of a Bayesian jury, which is characterized by Proposition 14. The Bayesian jury computes the probability that the defendant is guilty after observing $\hat{k}$ reports as a a function of the strategies of true and false victims. The share of potential false victims ($\gamma$) choosing to report is labeled $q_f$.

**Proposition 14.** *A Bayesian jury observing $\hat{k}$ reports chooses to convict the defendant if and only if $\hat{k} \geq k^{\star}(q, q_f, \overline{p})$.*

Proposition 14 shows that there will exist a minimum threshold of reports, which is given by $k^{\star}(q, q_f, \overline{p})$. Intuitively, a higher $\overline{p}$, which is the burden of proof, translates into a larger threshold of reports for conviction. The intuition for why the necessary number of reports for conviction depends on $q$ is easiest to explain when fixing the level of false reporting. A Bayesian jury observes $\hat{k}$ reports, and weighs the probability of observing $\hat{k}$ reports when the defendant is guilty against the probability of observing $\hat{k}$ reports when the defendant is innocent. If $q$ is very low, the jury knows that few true victims report, which means that observing more than one report against a guilty defendant is relatively less likely for a given value of $q_f$. However, as

$q$ increases, it becomes relatively more likely to observe more reports against a guilty offender. Theoretically, $k^\star(q, q_f, \overline{p})$ may also *increase* in $q$ for certain distributions of crime.[24]

Proposition 14 shows that a Bayesian jury may amplify the spillover effects from the previous section. More reporting (a larger $q$) will not only increase the probability of observing more than a certain number of reports, but more reporting may also affect the necessary number of reports for conviction.

### 3.4.2 Analysis of reporting decisions

Given the information held by a true victim, the probability that there will be enough reports for conviction is given by $p^v(q, q_f) = P^v[k \geq k^\star(q, q_f, \overline{p})|q, q_f]$. The level of reporting enters into this probability two times; $q$ affects both the probability of reaching a certain number of reports and the number of reports necessary for conviction. False victims can also compute the probability that a report leads to conviction given their information set, which is given by $p^f(q, q_f) = P^f[k \geq k^\star(q, q_f, \overline{p})|q, q_f]$.

**Proposition 15.** *In a Perfect Bayesian Equilibrium, a Bayesian jury convicts the defendant after observing at least $k^\star(q, q_f, \overline{p})$ reports. The reporting strategy of true victims is given by*

$$r_v(c_i) = \begin{cases} 1 & \textit{if } c_i \leq \hat{c}, \\ 0 & \textit{if } c_i > \hat{c}. \end{cases}$$

[24] Suppose that there are very few false victims and that most criminals are guilty of more than one crime. When relatively few true victims report, one report may be a strong signal of guilt. However, when more victims report it becomes relatively less likely to observe one report against a guilty defendant. This may imply that criminals that are guilty of one crime will not be convicted when a larger share of victims are reporting the crime.

*The reporting strategy of false victims is given by*

$$r_f(c_i) = \begin{cases} 1 & \textit{if } c_i \leq \hat{c}_f, \\ 0 & \textit{if } c_i > \hat{c}_f. \end{cases}$$

*The thresholds $\hat{c}$ and $\hat{c}_f$ are given by $\hat{c} = F^{-1}(q^\star)$ and $\hat{c}_f = F^{-1}(q_f^\star)$, where $q^\star = F[bp^v(q^\star, q_f^\star)]$ and $q_f^\star = F[bp^f(q^\star, q_f^\star)]$.*

There is an equilibrium where both true and false victims make a report if the costs of reporting is below a certain threshold, which means that Proposition 15 is a natural extension of Proposition 11. Generally, the conviction decision of a Bayesian jury may involve cases of jump discontinuity, which makes the equilibrium analysis more complicated. In this paper, I am primarily interested in analyzing the effect of different welfare-increasing mechanisms as well as small changes in reporting. In these cases, the potential discontinuities will not affect the analysis.

### 3.4.3 Social welfare

Section 3.3 shows that a small increase in crime reporting in equilibrium increases social welfare. In this section, the effect from a small increase in reporting on the strategies of other agents must also be taken into account. A small increase in reporting changes social welfare by

$$\frac{\partial W_{r_v(c_i)}}{\partial q} = \Pi[b\frac{\partial p^v(q^\star, q_f^\star)}{\partial q} q^\star] + \frac{\partial s_w()}{\partial q} C_w. \tag{3.3}$$

The last term in Equation (3.3) is the equilibrium effect of more true reporting on the probability of being wrongfully convicted.[25] The following result then follows as a corollary from Proposition 12.[26]

**Corollary 7.** *Suppose there is an equilibrium where $q^\star > 0$. Consider a reporting strategy given by*

$$r_v^\star(c_i) = \begin{cases} 1 & \textit{if } c_i \leq \hat{c} + \Delta, \\ 0 & \textit{if } c_i > \hat{c} + \Delta. \end{cases}$$

*For any small $\Delta$, $W_{r_v^\star(c_i)} \geq W_{r_v(c_i)}$ if $\frac{\partial W_{r_v(c_i)}}{\partial q} \geq 0$.*

As long as more truthful reporting does not lead to a sufficiently large spillover effect to false reporting, then the insights from Section 3.3 can be translated directly. A positive externality occurs because more reporting will increase the probability that other victims are able to convict their offenders.

### 3.4.4 Mechanisms to increase reporting and social welfare

Here, I analyze if it is possible to construct mechanisms to increase social welfare in the general model.

#### Adjusting the burden of proof and size of the punishment

In Section 3.3, the optimal burden of proof involves a trade-off between convicting criminals and avoiding wrongful convictions. The same intuition applies in this general model. The social planner can adjust $\overline{p}$, which influences $k^\star(q, q_f, \overline{p})$, thereby

[25] A small increase in truthful reporting implies that it becomes more likely for false victims to make a report against an agent that will be convicted, which leads to more incentives for false reporting.

[26] A small increase in reporting will not affect the number of reports necessary for conviction.

indirectly affecting the probability of conviction. If the social planner sets a low $\overline{p}$, there will be more truthful reporting, but also more wrongful convictions.

**Information Escrow**

The Information Escrow is optimal given that only true victims respond to the incentives created by the mechanism, but this mechanism may lead to an outcome that is very different from the social optimum with endogenous false reporting. The reason is that false victims respond to the same incentives as real victims, which means that promising that the case only goes to court when there is sufficient evidence will attract more false reports. A Bayesian jury is then unable to get precise signals of guilt, which leads to few convictions. Section 3.5.2 provides an example where the Information Escrow leads to an uninformative cheap talk equilibrium. For certain parameter values, I will show that a Bayesian jury needs to observe at least 24 reports to be sufficiently certain of guilt to make a conviction, which means that convictions will not occur in equilibrium.

### 3.4.5 Optimal mechanism with a Bayesian jury and strategic false victims

An optimal mechanism maximizes the welfare of victims and minimizes the number of wrongful convictions. Discouraging false reporting does not only lower the number of wrongful convictions, but also increases the informativeness of a report from a true victim. A difference between true and false victims is that true victims know that their offender is guilty, while the false victims do not receive any such signal. This is also the *only* difference between true and false victims in this model.

A separating equilibrium, where only the true victims report, needs to exploit differences in the conviction probability. Introducing a fee, $t_r$, conditional on non-conviction, may discourage false reporting, and will have a larger effect on false reporting when $p^f() \leq p^v()$. Only setting a large $t_r$ does not lead to a separating equilibrium. Suppose there is a separating equilibrium where no false victims report. This means that $\hat{k} = 1$ report is enough for a Bayesian jury to make a conviction, such that $p^f() = 1$, which means that true and false victims get the same benefits from reporting. Hence, it is necessary that $p^f() < 1$ to implement a separating equilibrium.

I construct a mechanism where all victims pay a cost $t_r$ of taking the case to court. This fee is returned to victims conditional on winning in court, such that budget-balance is satisfied.[27] If the jury observes only one report, the jury chooses *not* to convict with a probability $1 - r$ also in cases where the posterior probability exceeds the threshold for guilt. Suppose there is a separating equilibrium using this mechanism. By letting $P(k = 1)$ denote the probability for true victims that there are no other victims of the same criminal, the conviction probability for true victims is given by $p^v(r) = P(k = 1)r + [1 - P(k = 1)]$ in a separating equilibrium.[28] If false victims report, the conviction probability is $p^f(r) = (1 - \pi)r + \pi$. It is possible to sustain a separating equilibrium as long as the probability of making a report against a criminal with more than one crime $(1 - P(k = 1))$, is larger than the share of criminals $(\pi)$. As the share of criminals typically is small, while at least some criminals often are guilty of more than one crime, it is not unreasonable to assume that $\pi < 1 - P(k = 1)$.

**Definition 2.** *A pledging mechanism consists of each victim paying a fee $t_r$ to take*

[27] Budget-balance here implies that if the victims pay $t_r$ and win with probability $q$, then the returned amount is $\frac{t_r}{q}$.

[28] Here $P(k = 1)$ is *not* the share of criminals guilty of one crime. It is more likely to be the victim of a criminal with more crimes.

*the case to court. If the defendant is convicted, the fee is returned to the victim such that budget-balance is satisfied. If one report is observed, the jury chooses not to convict with a probability* $1 - r$.

**Proposition 16.** *Suppose* $t_r$ *can be arbitrarily large. If* $\pi < 1 - P(k = 1)$*, the pledging mechanism will implement an outcome where* $q \to 1$*,* $q_f \to 0$ *and* $p^v(r) \to 1$*, which will maximize social welfare.*

Proposition 16 is intuitive. A true victim will get the pledge returned with a larger probability using this mechanism, so a separating equilibrium will exist when the size of the pledge is sufficiently large. In the limit, this separating equilibrium implements an outcome where there are no wrongful convictions and all true victims convict their offender with a probability close to one. The mechanism implements the optimal outcome for each true victim and zero wrongful convictions, which means that the mechanism maximizes social welfare.

### Restricting $t_r$

In practice, there is an upper bound on the pledging fee, which I label $\bar{t}_r$. This means that a social planner cannot simultaneously implement a separating equilibrium and a conviction probability arbitrarily close to one for true victims. Suppose the social planner follows the Blackstone's ratio, by putting a large weight on avoiding wrongful convictions (Volokh, 1997). To implement an equilibrium without false reporting, it is necessary that

$$p^f(r)[b + \frac{\bar{t}_r}{p^v(r)}] - \underline{c} - \bar{t}_r < 0. \tag{3.4}$$

Equation (3.4) is not necessarily satisfied for $r$ close to $1$ when there is an upper bound for the pledging fee. In this case, decreasing the probability of conviction after one report, which corresponds to setting a lower $r$, may be necessary to discourage false reporting. A lower value of $r$ may also discourage true victims from reporting, but there is a separating equilibrium where all true victims report if

$$p^v(r)[b + \frac{\bar{t}_r}{p^v(r)}] - \bar{c} - \bar{t}_r = p^v(r)b - \bar{c} \geq 0. \tag{3.5}$$

If a sufficiently low $r$ is necessary to discourage false reporting, then Equation (3.5) is not satisfied. In this case, a social planner that wants to minimize wrongful convictions can implement a semi-separating equilibrium where some true victims choose not to report. In the extreme case where the welfare weight on avoiding wrongful convictions is sufficiently large and $\bar{t}_r$ is small, then a social planner may optimally implement an outcome with a low level of reporting from true victims and a low conviction probability.

## 3.5 Numerical example

I construct a numerical example where half of the criminals are guilty of one crime, while the other half are guilty of two crimes. The share of criminals is given by $\pi = 0.05$. The distribution of the costs of reporting is given by $c_i \sim U[0, 1]$. I set $b = 1.1$, which means that the cost of taking the case to court is smaller than the benefits from conviction for all victims. A share $\gamma = 0.005$ are potential false victims.

### 3.5.1 Example from the basic model

In the version of the example from the basic model, I set $\tilde{k} = 2$ reports as the threshold for conviction, and I let the level of false reporting be given by $\gamma = 0.005$. A report from a true victim leads to a conviction in two cases. There may be at least one false report, which happens with probability $1 - e^{-0.005} \approx 0.005$. Alternatively, there may be another victim of the same criminal reporting, which happens with probability $\frac{2q}{3}$ when a share $q$ of victims report.[29] This leads to a conviction probability of $p^v(q) = 0.005 + 0.995\frac{2q}{3}$.

#### Equilibrium

The equilibrium is given by victims reporting if $c_i \leq bp^v(q^\star)$. Here $q^\star$ solves $q^\star = bp^v(q^\star)$, which leads to $q^\star = 1.1[0.005 + 0.995\frac{2q^\star}{3}]$ or $q^\star \approx 0.02$. The probability of conviction is given by $p^v(q^\star) = 0.018$. The cost of taking the case to court is strictly lower than the benefits of conviction for all victims, but still the reporting rate is close to zero in equilibrium.

**Comparative statics** Small changes in parameter values may have large equilibrium effects because of the spillover effects associated with reporting. Consider a small increase from $b = 1.1$ to $b_2 = 1.2$. The direct effect of a marginal increase in $b$ is given by $(b_2 - b)p^v(q^\star) = 0.0018$, but the equilibrium effect will multiply the direct effect with a factor $\frac{1}{1-bF'(\cdot)\frac{\partial p^v(q^\star)}{\partial q}}$. For the parameter values in this example, the equilibrium effect is roughly four times larger than the direct effect. A small change in $b$ may hence have a large effect on reporting.

[29] Although criminals are equally likely to have one and two victims, the probability of being the victim of a criminal guilty of two crimes is $\frac{2}{3}$ because this half of criminals will have two victims.

**Social welfare** Social welfare depends on the welfare of victims of crime and on the number of wrongful convictions. When $\tilde{k} = 2$ reports are necessary for convictions, approximately zero wrongful convictions will be made given that $\gamma = 0.005$.[30] The social welfare in equilibrium is given by

$$W(q^\star) = \int_0^{q^\star} [bp^v(q^\star) - c_i]dc_i = \frac{[bp^v(q^\star)]^2}{2} = \frac{(1.1 * 0.018)^2}{2} \approx 0.$$

There are almost no victims reporting, and the reporting victims only win in court with a small probability. If all victims report, the conviction probability is given by $p^v(1) = 0.005 + 0.995 * \frac{2}{3} = 0.67$, and the social welfare is given by $W(1) = \int_0^1 [bp^v(1) - c_i]dc_i = \frac{(1.1*0.67)^2}{2} = 0.27$.[31] This outcome is not an equilibrium. The expected benefits when reporting are $bp^v(1) \approx 0.74$ in this case, which implies that roughly one quarter of victims would choose not to report given that everyone else reports. Given that these victims choose not to report, the expected benefits of reporting are $bp^v(0.74) \approx 0.55$, which implies that victims with costs between $0.55$ and $0.74$ also will choose not to report. Iterating this argument leads to the equilibrium outcome.

**Mechanisms to increase social welfare**

**Adjusting burden of proof** By adjusting the burden of proof to $\tilde{k} = 1$, all true victims know that the probability of conviction after a report is $p^v(q) = 1$, which implies that all victims choose to report. But the probability of being wrongfully convicted increases to $s_w(\tilde{k} = 1) = 1 - e^{-0.005} \approx 0.005$, which implies that a trade-off occurs between increasing truthful reporting and avoiding wrongful convictions.

[30]The probability that each innocent agent is convicted is $1 - e^{-0.005} - 0.005e^{-0.005} \approx 0$.

[31]As described by Proposition 12, a marginal increase in reporting will increase social welfare by $b\frac{\partial p^v(q^\star)}{\partial q^\star}q^\star$. Taking the derivative of $W(q^\star) = \frac{[bp^v(q)]^2}{2}$ leads to the same expression.

The social welfare is given by

$$W_{\tilde{k}=1} = \int_0^1 (b - c_i)dc_i - s_w(\tilde{k} = 1)C_w = 0.6 - 0.005C_w.$$

Depending on the cost of wrongful convictions, $C_w$, lowering the burden of proof may increase or decrease social welfare.

**Payment** A payment given by $1 - bp^v(1) = 1 - 1.1 * 0.67 = 0.26$ leads to an outcome where all victims choose to report. I have assumed that there are no efficiency losses from such payments, such that the social welfare increases to $W(1) = 0.27$. While the payment clearly increases social welfare relative to the equilibrium outcome, this cannot be the optimal outcome. One third of the victims pay the cost of going to court without getting the benefits from conviction. The information asymmetry is not being resolved by the payments, because these victims would be better off not reporting.

**Information Escrow** The Information Escrow credibly promises the victim that the case is taken to court if and only if there is sufficient evidence for conviction. All victims report a crime to the principal using this mechanism. A conviction will be made if another agent reports against the same offender, which happens with probability $p^v(1) = 0.667$. In this case, the victims get benefits $b = 1.1$ and pay a cost $c_i \leq 1$. With probability $1 - p^v(1) = 0.33$, the victims get zero benefits and zero costs. Given $\tilde{k} = 2$, there are approximately zero wrongful convictions, which means that the social welfare is given by

$$W_{InfEsc} = \int_0^1 p^v(1)(b - c_i)dc_i = 0.4.$$

The social welfare using the Information Escrow exceeds the equilibrium social welfare for given choices of $\tilde{k}$.[32]

### 3.5.2 Example from the general model

I now use the same example to analyze the outcome under the different mechanisms in the general model.

#### Information Escrow

Using the Information Escrow, the equilibrium is given by all true victims as well as all $\gamma$ false victims reporting to the principal. The conviction decision depends on the threshold $\overline{p}$. I here assume that $\overline{p} = 0.95$, which implies that the threshold for conviction is given by $k^{\star}(q = 1, q_f = 1, \overline{p}) = 2$. In this example, there are few false crime reports, which means that observing 2 reports is a strong signal of guilt.

**Varying the size of $\gamma$** I here show that varying the size of $\gamma$ has a large effect on the Bayesian jury. In the limit as $\gamma \to 1$, then at least $24$ reports are necessary for conviction using the Information Escrow, which means that the outcome is given by an uninformative cheap talk equilibrium. Letting $\gamma \to 1$ is an extreme case, but similar effects are observed for lower values of $\gamma$. As long as $\gamma > 0.11$, at least $4$ reports are necessary for conviction, which means that the probability of conviction is low for all criminals.[33]

[32]For $\tilde{k} = 1$ the social welfare is similar. The reason is that setting $\tilde{k} = 1$ eliminates all issues of information asymmetry. For larger values of $\tilde{k}$, the social welfare using the Information Escrow is larger than the equilibrium outcome.

[33]When $\gamma = 0.11$, the probability of conviction for a criminal guilty of two crimes is given by $1 - \sum_{i=0}^{i=1} e^{-0.11} \frac{0.11^i}{i!} = 0.005$.

**Pledging mechanism**

Here I analyze if the pledging mechanism can be used to construct a separating equilibrium, where all true victims and no false victims report a crime. In a separating equilibrium, the true victims get the pledge returned with certainty if another victims also reports (which happens with probability $\frac{2}{3}$) and with probability $r$ otherwise. This leads to a winning probability of $p^v(r) = \frac{2}{3} + \frac{r}{3}$. For budget-balance to be satisfied, it is necessary that winning victims are paid back $\frac{3t_r}{2+r}$.[34]

If false victims deviate from the separating equilibrium and choose to report, this leads to a conviction with probability $p^f(r) = 0.05 + 0.95r$, because a report is directed towards a criminal with probability $\pi = 0.05$. The false victims with lowest cost of reporting ($c_i = 0$) will choose not to report as long as $(0.05 + 0.95r)(\frac{3t_r}{2+r} + 1.1) - t_r \leq 0$. If any $t_r$ is feasible, then setting $r$ arbitrarily close to 1 will discourage false reporting for large values of $t_r$. Setting $t_r = 1000$ implements a separating equilibrium for $r = 0.99$.

**Social welfare** All true victims report and get $b - c_i$, while there are no wrongful convictions. Hence the social welfare by setting a large $t_r$ and $r$ close to one is given by

$$W_{pl} = \lim_{r \to 1} \int_0^1 p^v(r)(b - c_i)dc_i = 0.6.$$

The main contribution of the pledging mechanism is to discourage false reporting, which has two positive welfare effects. The number of wrongful convictions is reduced to zero, while all true victims win in court with a high probability because the Bayesian jury gets a precise signal of guilt. The same effect occurs for other values of $\gamma$. Using the information Escrow, there is only an uninformative equilibrium

[34] The victims pay $t_r$ and win with probability $\frac{2}{3} + \frac{r}{3}$.

when $\gamma = 0.2$, but the pledging mechanism also implements a separating equilibrium for higher values of $\gamma$.

**Restricting** $t_r$ I now assume the maximum size of the pledge is $\bar{t}_r = 5$. To discourage false reporting the probability of winning after one report ($r$) must be reduced, and setting $r = 0.75$ leads to $p^v(r = 0.75) = 0.92$ and $p^f(r = 0.75) = 0.76$. Lowering the probability of conviction after one report to $r = 0.75$ leads to a separating equilibrium. The reason is that this reduction will have a larger impact for false victims, because an outcome where one report is observed is more likely for these agents.

## 3.6 Conclusion

This paper analyzes the incentives for crime reporting. Two central insights emerge from this paper. The first insight is that under-reporting of crime generally occurs in equilibrium, and the second insight is that the effect of measures to incentivize crime reporting depends on the behavior of potential false victims. While a social planner always wants to increase crime reporting, it may be impossible to create mechanisms that only incentivize true reporting. In circumstances where false reporting is a minor issue, the Information Escrow will increase crime reporting and social welfare. Empirical evidence shows low levels of false reporting, which suggests that the Information Escrow will increase social welfare. This paper supports this argument in certain settings, but I also argue that the relation may be more complicated. Low observed levels of false reporting can also be understood as the outcome of a judicial system that discourages false reporting by making reporting costly and uncertain.

# References

**Abadie, Alberto, Alexis Diamond, and Jens Hainmueller**, "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American statistical Association*, 2010, *105* (490), 493–505.

_ **and Javier Gardeazabal**, "The economic costs of conflict: A case study of the Basque Country," *American Economic Review*, 2003, *93* (1), 113–132.

**Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin**, "A political theory of populism," *The Quarterly Journal of Economics*, 2013, *128* (2), 771–805.

**Akkerman, Tjitske**, "Comparing radical right parties in government: Immigration and integration policies in nine countries (1996–2010)," *West European Politics*, 2012, *35* (3), 511–529.

**Al-Ubaydli, Omar, John A List, Danielle LoRe, and Dana Suskind**, "Scaling for Economists: Lessons from the non-adherence problem in the medical literature," *Journal of Economic Perspectives*, 2017, *31* (4), 125–44.

**Algan, Yann, Sergei Guriev, Elias Papaioannou, and Evgenia Passari**, "The European trust crisis and the rise of populism," *Brookings Papers on Economic Activity*, 2017, *2017* (2), 309–400.

**Allen, W David**, "The reporting and underreporting of rape," *Southern Economic Journal*, 2007, pp. 623–641.

**Anesi, Vincent and Philippe De Donder**, "Party formation and minority ideological positions," *The Economic Journal*, 2009, *119* (540), 1303–1323.

**Austen-Smith, David and Michael Wallerstein**, "Redistribution and affirmative action," *Journal of Public Economics*, 2006, *90* (10-11), 1789–1823.

**Autor, David, David Dorn, Gordon Hanson, Kaveh Majlesi et al.**, *Importing political polarization? The electoral consequences of rising trade exposure* number w22637 2016.

**Ayres, Ian and Cait Unkovic**, "Information Escrows," *Michigan Law Review*, 2012, *111*, 145.

_ **and Steven D Levitt**, "Measuring positive externalities from unobservable victim precaution: an empirical analysis of Lojack," *The Quarterly Journal of Economics*, 1998, *113* (1), 43–77.

**Bagnoli, Mark and Barton L Lipman**, "Provision of public goods: Fully implementing the core through private contributions," *The Review of Economic Studies*, 1989, *56* (4), 583–601.

**Balkin, Steven and John F McDonald**, "The market for street crime: An economic analysis of victim-offender interaction," *Journal of Urban Economics*, 1981, *10* (3), 390–405.

**Bartels, Larry M et al.**, "What's the Matter with What's the Matter with Kansas?," *Quarterly Journal of Political Science*, 2006, *1* (2), 201–226.

**Barth, Erling, Henning Finseraas, and Karl O Moene**, "Political reinforcement: how rising inequality curbs manifested welfare generosity," *American Journal of Political Science*, 2015, *59* (3), 565–577.

**Becker, Gary S**, "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 1968, *76* (2), 169–217.

**Benabou, Roland**, "Unequal societies: Income distribution and the social contract," *American Economic Review*, 2000, *90* (1), 96–129.

**_ and Efe A Ok**, "Social mobility and the demand for redistribution: the POUM hypothesis," *The Quarterly Journal of Economics*, 2001, *116* (2), 447–487.

**Benartzi, Shlomo, John Beshears, Katherine L Milkman, Cass R Sunstein, Richard H Thaler, Maya Shankar, Will Tucker-Ray, William J Congdon, and Steven Galing**, "Should governments invest more in nudging?," *Psychological Science*, 2017, *28* (8), 1041–1055.

**Bertrand, Marianne and Emir Kamenica**, "Coming apart? Cultural distances in the United States over time," Technical Report, National Bureau of Economic Research 2018.

**Bhatti, Yosef, Jens Olav Dahlgaard, Jonas Hedegaard Hansen, and Kasper M Hansen**, "How voter mobilization from short text messages travels within households and families: Evidence from two nationwide field experiments," *Electoral Studies*, 2017, *50*, 39–49.

**_ , _ , _ , and _** , "Moving the campaign from the front door to the front pocket: field experimental evidence on the effect of phrasing and timing of text messages on voter turnout," *Journal of Elections, Public Opinion and Parties*, 2017, *27* (3), 291–310.

**Black, Duncan**, "On the rationale of group decision-making," *Journal of Political Economy*, 1948, *56* (1), 23–34.

**Bó, Ernesto Dal, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne**, "Economic Losers and Political Winners: Sweden's Radical Right," 2019.

**Bond, Robert M, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler**, "A 61-million-person experiment in social influence and political mobilization," *Nature*, 2012, *489* (7415), 295.

**Bonica, Adam, Nolan McCarty, Keith T Poole, and Howard Rosenthal**, "Why hasn't democracy slowed rising inequality?," *Journal of Economic Perspectives*, 2013, *27* (3), 103–24.

**Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman**, "Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions," *Econometrica*, 2014, *82* (4), 1273–1301.

**Calvert, Randall L**, "Robustness of the multidimensional voting model: Candidate motivations, uncertainty, and convergence," *American Journal of Political Science*, 1985, pp. 69–95.

**Carlsson, Hans and Eric Van Damme**, "Global games and equilibrium selection," *Econometrica*, 1993, pp. 989–1018.

**Carrell, Scott E, Richard L Fullerton, and James E West**, "Does your cohort matter? Measuring peer effects in college achievement," *Journal of Labor Economics*, 2009, *27* (3), 439–464.

**Chalfin, Aaron and Justin McCrary**, "Criminal deterrence: A review of the literature," *Journal of Economic Literature*, 2017, *55* (1), 5–48.

**Chassang, Sylvain and Gerard Padró i Miquel**, "Corruption, Intimidation, and Whistleblowing: A Theory of Inference from Unverifiable Reports," *The Review of Economic Studies*, 2018.

**Cheng, Ing-Haw and Alice Hsiaw**, "Reporting Sexual Misconduct in the# MeToo Era," *Working Paper*, 2019.

**Cook, Philip J**, "The Demand and Supply of Criminal Opportunities," *Crime & Justice*, 1986, *7*, 1.

_ **and John MacDonald**, "Public safety through private action: an economic assessment of BIDS," *The Economic Journal*, 2011, *121* (552), 445–462.

**Cox, DR**, "Some Problems Connected with Statistical Inference," *The Annals of Mathematical Statistics*, 1958, *29* (2), 357–372.

**Dahl, Gordon B, Katrine V Løken, and Magne Mogstad**, "Peer effects in program participation," *American Economic Review*, 2014, *104* (7), 2049–74.

**Dale, Allison and Aaron Strauss**, "Don't forget to vote: text message reminders as a mobilization tool," *American Journal of Political Science*, 2009, *53* (4), 787–804.

**Davis, Jonathan, Jonathan Guryan, Kelly Hallberg, and Jens Ludwig**, "The Economics of Scale-Up," Technical Report, National Bureau of Economic Research 2017.

**Deaton, Angus and Nancy Cartwright**, "Understanding and misunderstanding randomized controlled trials," *Social Science & Medicine*, 2018, *210*, 2–21.

**DellaVigna, Stefano, John A List, Ulrike Malmendier, and Gautam Rao**, "Voting to tell others," *The Review of Economic Studies*, 2016, *84* (1), 143–181.

**Dennison, James and Teresa Talò**, "Explaining attitudes to immigration in France," 2017.

**Desmet, Klaus and Romain Wacziarg**, "The Cultural Divide," Technical Report, National Bureau of Economic Research 2018.

**Donder, P De and M Gallego**, "Electoral Competition and Party Positioning," Technical Report, TSE Working Paper, 17-760 2017.

**Downs, A.**, *An Economic Theory of Democracy*, Harper, 1957.

**Draca, Mirko and Stephen Machin**, "Crime and Economic Incentives," *Annual Review of Economics*, 2015, *7*, 389–408.

**Duca, John V and Jason L Saving**, "Income inequality and political polarization: time series evidence over nine decades," *Review of Income and Wealth*, 2016, *62* (3), 445–466.

**Duflo, Esther and Emmanuel Saez**, "The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment," *The Quarterly journal of economics*, 2003, *118* (3), 815–842.

**Duggan, John**, "A survey of equilibrium analysis in spatial models of elections," 2005.

— **and Matthew Jackson**, "Mixed strategy equilibrium and deep covering in multidimensional electoral competition," *Typescript, University of Rochester*, 2005.

**Dziubiński, Marcin and Jaideep Roy**, "Electoral competition in 2-dimensional ideology space with unidimensional commitment," *Social Choice and Welfare*, 2011, *36* (1), 1–24.

**Edsall, Thomas Byrne and Mary D Edsall**, *Chain reaction: The impact of race, rights, and taxes on American politics*, WW Norton & Company, 1992.

**Egorov, Georgy**, "Single-issue campaigns and multidimensional politics," Technical Report, National Bureau of Economic Research 2015.

**Ehrlich, Isaac**, "Participation in illegitimate activities: A theoretical and empirical investigation," *Journal of Political Economy*, 1973, *81* (3), 521–565.

— , "Crime, punishment, and the market for offenses," *Journal of Economic Perspectives*, 1996, *10* (1), 43–67.

**Enke, Benjamin**, "Moral values and voting," Technical Report, National Bureau of Economic Research 2018.

**Fernández, Raquel and Gilat Levy**, "Diversity and redistribution," *Journal of Public Economics*, 2008, *92* (5-6), 925–943.

**Fiorina, Morris P and Samuel J Abrams**, "Political polarization in the American public," *Annual Review of Political Science*, 2008, *11*, 563–588.

**Fisher, Bonnie S, Leah E Daigle, Francis T Cullen, and Michael G Turner**, "Reporting sexual victimization to the police and others: Results from a national-level study of college women," *Criminal Justice and Behavior*, 2003, *30* (1), 6–38.

**Fowler, Anthony**, "Electoral and policy consequences of voter turnout: Evidence from compulsory voting in Australia," 2013.

**Frank, Thomas**, *What's the Matter with Kansas?*, Metropolitan Books, 2004.

**Fredriksson, Per G, Le Wang, and Khawaja A Mamun**, "Are politicians office or policy motivated? The case of US governors' environmental policies," *Journal of Environmental Economics and Management*, 2011, *62* (2), 241–253.

**Furlong, William J**, "A general equilibrium model of crime commission and prevention," *Journal of Public Economics*, 1987, *34* (1), 87–103.

**Gerber, Alan S and Donald P Green**, "Field Experiments on Voter Mobilization: An Overview of a Burgeoning Literature," in "Handbook of Economic Field Experiments," Vol. 1, Elsevier, 2017, pp. 395–438.

_ **and Todd Rogers**, "Descriptive social norms and motivation to vote: Everybody's voting and so should you," *The Journal of Politics*, 2009, *71* (1), 178–191.

— , **Donald P Green, and Christopher W Larimer**, "Social pressure and voter turnout: Evidence from a large-scale field experiment," *American Political Science Review*, 2008, *102* (1), 33–48.

**Glaeser, Edward L, Giacomo AM Ponzetto, and Jesse M Shapiro**, "Strategic extremism: Why Republicans and Democrats divide on religious values," *The Quarterly Journal of Economics*, 2005, *120* (4), 1283–1330.

**Goldberg, Itzhak and Frederick C Nold**, "Does Reporting Deter Burglars?–An Empirical Analysis of Risk and Return in Crime," *The Review of Economics and Statistics*, 1980, pp. 424–431.

**Golder, Matt**, "Far right parties in Europe," *Annual Review of Political Science*, 2016, *19*, 477–497.

**Green, Donald P and Alan S Gerber**, *Get out the vote: How to increase voter turnout*, Brookings Institution Press, 2015.

— , **Anna Wilke, and Jasper Cooper**, "Countering violence against women at scale: A mass media experiment in rural Uganda," *Working Paper*, 2019.

— , **Mary C McGrath, and Peter M Aronow**, "Field experiments and the study of voter turnout," *Journal of Elections, Public Opinion & Parties*, 2013, *23* (1), 27–48.

**Guiso, Luigi, Helios Herrera, Massimo Morelli, and Tommaso Sonno**, "Populism: Demand and Supply," 2017.

**Hausman, Catherine and David S Rapson**, "Regression discontinuity in time: Considerations for empirical applications," *Annual Review of Resource Economics*, 2018, *10*, 533–552.

**Heath, Anthony and Lindsay Richards**, "Attitudes Towards Immigration and their Antecedents: Topline Results from Round 7 of the European Social Survey," *ESS Topline Results Series*, 2016, *7.*

**Heckman, James J**, "The scientific model of causality," *Sociological Methodology*, 2005, *35* (1), 1–97.

**Hill, Kim Quaile, Jan E Leighley, and Angela Hinton-Andersson**, "Lower-class mobilization and policy linkage in the US states," *American Journal of Political Science*, 1995, pp. 75–86.

**Holland, Paul W**, "Statistics and causal inference," *Journal of the American Statistical Association*, 1986, *81* (396), 945–960.

**Ivarsflaten, Elisabeth**, "What unites right-wing populists in Western Europe? Re-examining grievance mobilization models in seven successful cases," *Comparative Political Studies*, 2008, *41* (1), 3–23.

**Iversen, Torben and Max Goplerud**, "Redistribution without a median voter: Models of multidimensional politics," *Annual Review of Political Science*, 2018, *21*, 295–317.

**Iyer, Lakshmi, Anandi Mani, Prachi Mishra, and Petia Topalova**, "The power of political voice: women's political representation and crime in India," *American Economic Journal: Applied Economics*, 2012, *4* (4), 165–93.

**Jones, Jason J, Robert M Bond, Eytan Bakshy, Dean Eckles, and James H Fowler**, "Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 US presidential election," *PloS one*, 2017, *12* (4), e0173851.

**Karabarbounis, Loukas**, "One dollar, one vote," *The Economic Journal*, 2011, *121* (553), 621–651.

**Karadja, Mounir, Johanna Mollerstrom, and David Seim**, "Richer (and holier) than thou? The effect of relative income improvements on demand for redistribution," *Review of Economics and Statistics*, 2017, *99* (2), 201–212.

**Karakostas, Alexandros and Daniel John Zizzo**, "Compliance and the power of authority," *Journal of Economic Behavior & Organization*, 2016, *124*, 67–80.

**Krasa, Stefan and Mattias K Polborn**, "Political competition between differentiated candidates," *Games and Economic Behavior*, 2012, *76* (1), 249–271.

_ **and Mattias Polborn**, "Social ideology and taxes in a differentiated candidates framework," *American Economic Review*, 2014, *104* (1), 308–22.

**Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva**, "How elastic are preferences for redistribution? Evidence from randomized survey experiments," *American Economic Review*, 2015, *105* (4), 1478–1508.

**Lee, David S and Thomas Lemieux**, "Regression discontinuity designs in economics," *Journal of economic literature*, 2010, *48* (2), 281–355.

**Lee, Frances Xu and Wing Suen**, "Credibility of crime allegations," *American Economic Journal-Microeconomics*, 2019.

**Lee, Woojin and John E Roemer**, "Racism and redistribution in the United States: A solution to the problem of American exceptionalism," *Journal of Public Economics*, 2006, *90* (6-7), 1027–1052.

**Levy, Ro'ee and Martin Mattson**, "The Effects of Social Movements: Evidence from #MeToo," *Working Paper*, 2019.

**Lijphart, Arend**, "Unequal participation: Democracy's unresolved dilemma presidential address, American Political Science Association, 1996," *American Political Science Review*, 1997, *91* (1), 1–14.

__ , "The problem of low and unequal voter turnout-and what we can do about it," 1998.

__ , *Patterns of democracy: Government forms and performance in thirty-six countries*, Yale University Press, 2012.

**Lindbeck, Assar and Jörgen W Weibull**, "Balanced-budget redistribution as the outcome of political competition," *Public Choice*, 1987, *52* (3), 273–297.

**Lindqvist, Erik and Robert Östling**, "Identity and redistribution," *Public Choice*, 2013, *155* (3-4), 469–491.

**Lisak, David and Paul M Miller**, "Repeat rape and multiple offending among undetected rapists," *Violence and Victims*, 2002, *17* (1), 73.

**__ , Lori Gardinier, Sarah C Nicksa, and Ashley M Cote**, "False allegations of sexual assault: An analysis of ten years of reported cases," *Violence Against Women*, 2010, *16* (12), 1318–1334.

**López, Ian Haney**, *Dog whistle politics: How coded racial appeals have reinvented racism and wrecked the middle class*, Oxford University Press, 2015.

**Luce, Helen, Sarina Schrager, and Valerie Gilchrist**, "Sexual assault of women," *American Family Physician*, 2010, *81* (4), 489–495.

**Mahler, Vincent A**, "Electoral turnout and income redistribution by the state: A cross-national analysis of the developed democracies," *European Journal of Political Research*, 2008, *47* (2), 161–183.

**Martin, Lanny W and Randolph T Stevenson**, "Government formation in parliamentary democracies," *American Journal of Political Science*, 2001, pp. 33–50.

**Marx, Karl, Karl Marx, Friedrich Engels, and Lasker**, *Marx, Engels: Selected Correspondence*, Progress Publishers, 1975.

**Matakos, Konstantinos and Dimitrios Xefteris**, "Divide and rule: redistribution in a model with differentiated candidates," *Economic Theory*, 2017, *63* (4), 867–902.

**McCarty, Nolan, Keith T Poole, and Howard Rosenthal**, *Polarized America: The dance of ideology and unequal riches*, MIT Press, 2016.

**McKelvey, Richard D**, "Intransitivities in multidimensional voting models and some implications for agenda control," *Journal of Economic Theory*, 1976, *12* (3), 472–482.

**Mello, Luiz De and Erwin R Tiongson**, "Income inequality and redistributive government spending," *Public Finance Review*, 2006, *34* (3), 282–305.

**Meltzer, Allan H and Scott F Richard**, "A rational theory of the size of government," *Journal of Political Economy*, 1981, *89* (5), 914–927.

**Mengeling, Michelle A, Brenda M Booth, James C Torner, and Anne G Sadler**, "Reporting sexual assault in the military: Who reports and why most servicewomen don't," *American Journal of Preventive Medicine*, 2014, *47* (1), 17–25.

**Morris, Stephen and Hyun Song Shin**, "Unique equilibrium in a model of self-fulfilling currency attacks," *American Economic Review*, 1998, pp. 587–597.

**Mudde, Cas**, *The ideology of the extreme right*, Manchester University Press, 2010.

_ **and Cristóbal Rovira Kaltwasser**, *Populism: A very short introduction*, Oxford University Press, 2017.

**Muralidharan, Karthik and Paul Niehaus**, "Experimentation at Scale," *Journal of Economic Perspectives*, 2017, *31* (4), 103–24.

**Nagin, Daniel S**, "Deterrence: A review of the evidence by a criminologist for economists," *Annual Review of Economics*, 2013, *5* (1), 83–105.

**Nickerson, David W**, "Is voting contagious? Evidence from two field experiments," *American Political Science Review*, 2008, *102* (1), 49–57.

**O, Ana L De La and Jonathan A Rodden**, "Does religion distract the poor? Income and issue voting around the world," *Comparative Political Studies*, 2008, *41* (4-5), 437–476.

**Palfrey, Thomas R and Howard Rosenthal**, "Participation and the provision of discrete public goods: a strategic analysis," *Journal of Public Economics*, 1984, *24* (2), 171–193.

**Panagopoulos, Costas**, "Timing is everything? Primacy and recency effects in voter mobilization campaigns," *Political Behavior*, 2011, *33* (1), 79–93.

**Paternoster, Raymond**, "How much do we really know about criminal deterrence?," *The Journal of Criminal Law and Criminology*, 2010, pp. 765–824.

**Patterson, Debra and Rebecca Campbell**, "Why rape survivors participate in the criminal justice system," *Journal of Community Psychology*, 2010, *38* (2), 191–205.

**Pei, Harry and Bruno Strulovici**, "Crime Entanglement, Deterrence and Accuser Credibility," *Working Paper*, 2019.

**Perez-Truglia, Ricardo and Guillermo Cruces**, "Partisan interactions: Evidence from a field experiment in the United States," *Journal of Political Economy*, 2017, *125* (4), 1208–1243.

**Peterson, Cora, Sarah DeGue, Curtis Florence, and Colby N Lokey**, "Lifetime economic burden of rape among US adults," *American Journal of Preventive Medicine*, 2017, *52* (6), 691–701.

**Piketty, Thomas**, "About capital in the twenty-first century," *American Economic Review*, 2015, *105* (5), 48–53.

_ **and Emmanuel Saez**, "Income inequality in the United States, 1913–1998," *The Quarterly Journal of Economics*, 2003, *118* (1), 1–41.

**Plott, Charles R**, "A notion of equilibrium and its possibility under majority rule," *The American Economic Review*, 1967, *57* (4), 787–806.

**Polborn, Mattias K and James M Snyder**, "Party polarization in legislatures with office-motivated candidates," *The Quarterly Journal of Economics*, 2017, *132* (3), 1509–1550.

**Poole, Keith T and Howard Rosenthal**, "Patterns of congressional voting," *American Journal of Political Science*, 1991, pp. 228–278.

**Quidt, Jonathan De, Johannes Haushofer, and Christopher Roth**, "Measuring and Bounding Experimenter Demand," Technical Report, National Bureau of Economic Research 2017.

**Roberts, Kevin WS**, "Voting over income tax schedules," *Journal of public Economics*, 1977, *8* (3), 329–340.

**Rodriguez, Francisco**, "Inequality, Redistribution, and Rent-seeking," *Economics & Politics*, 2004, *16* (3), 287–320.

**Roemer, John E**, "Why the poor do not expropriate the rich: an old argument in new garb," *Journal of Public Economics*, 1998, *70* (3), 399–424.

_ **and Karine Van der Straeten**, "Xenophobia and the size of the public sector in France: a politico-economic analysis," *Journal of Economics*, 2005, *86* (2), 95–144.

**Romer, Thomas**, "Individual welfare, majority voting, and the properties of a linear income tax," *Journal of Public Economics*, 1975, *4* (2), 163–185.

**Rosenbaum, Paul R**, "Interference between units in randomized experiments," *Journal of the American Statistical Association*, 2007, *102* (477), 191–200.

**Rotenberg, Cristine and Adam Cotter**, "Police-reported sexual assaults in Canada before and after# MeToo, 2016/2017," *Juristat*, 2018, *38* (1).

**Rubin, Donald B**, "Randomization analysis of experimental data: The Fisher randomization test comment," *Journal of the American Statistical Association*, 1980, *75* (371), 591–593.

**Rydgren, Jens**, "The sociology of the radical right," *Annual Review of Sociology*, 2007, *33*, 241–262.

**Sacerdote, Bruce**, "Peer effects with random assignment: Results for Dartmouth roommates," *The Quarterly journal of economics*, 2001, *116* (2), 681–704.

**Saez, Emmanuel and Gabriel Zucman**, *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*, W.W.Norton, 2019.

**Schelling, Thomas C**, *The strategy of conflict*, Harvard University Press, 1960.

**Shayo, Moses**, "A model of social identity with an application to political economy: Nation, class, and redistribution," *American Political Science Review*, 2009, *103* (2), 147–174.

**Sinclair, Betsy, Margaret McConnell, and Donald P Green**, "Detecting spillover effects: Design and analysis of multilevel experiments," *American Journal of Political Science*, 2012, *56* (4), 1055–1069.

**Spohn, Cassia, Clair White, and Katharine Tellis**, "Unfounding sexual assault: Examining the decision to unfound and identifying false reports," *Law & Society Review*, 2014, *48* (1), 161–192.

**Sunstein, Cass and Richard Thaler**, "Nudge: Improving decisions about health, wealth, and happiness," 2008.

**Tabarrok, Alexander**, "The private provision of public goods via dominant assurance contracts," *Public Choice*, 1998, *96* (3-4), 345–362.

**Tavits, Margit and Joshua D Potter**, "The effect of inequality and social identity on party strategies," *American Journal of Political Science*, 2015, *59* (3), 744–758.

**Verba, Sidney, Norman H Nie, and Jae O Kim**, "Participation and Political Equality: A Seven Nation Comparison," 1978.

**Volokh, Alexander**, "N guilty men," *University of Pennsylvania Law Review*, 1997, *146* (1), 173–216.

**Wittman, Donald**, "Candidates with policy preferences: A dynamic model," *Journal of Economic Theory*, 1977, *14* (1), 180–189.

**Wolfinger, Raymond E and Steven J Rosenstone**, *Who votes?*, Vol. 22, Yale University Press, 1980.

**Xefteris, Dimitrios**, "Multidimensional electoral competition between differentiated candidates," *Games and Economic Behavior*, 2017, *105*, 112–121.

## .1 Proofs for Section 1.3 (Analyzing the model)

**Lemma 1.** *If both parties propose the same cultural policy ($c_l = c_r$), then both parties will propose the median voter's preferred tax rate ($t_l = t_r = t^m(\beta)$).*

*Proof.* Suppose there is an equilibrium where party $j$ proposes $t^m(\beta)$ and party $-j$ propose $t_{-j} \neq t^m(\beta)$. In this case party $-j$ wins with probability $0$. Deviating to $t^m(\beta)$ leads to a winning probability of $\frac{1}{2}$, which will be a profitable deviation for $-j$. Hence there is no equilibrium where only one party proposes $t^m(\beta)$. Suppose there is an equilibrium where both parties propose different tax rates. If the median voter is indifferent, both parties win with probability $\frac{1}{2}$. An infinitesimally small deviation increases the winning probability to $1$ while only having a negligible effect on the policy in office, which means that this is a profitable deviation. If the median voter is not indifferent, then one party wins with certainty and I can repeat the argument from the case where only one party deviates from the median solution. □

**Lemma 2.** *The unique equilibrium in the subgame where $c_l = 0$ and $c_r = 1$ is given by ($t_l = \frac{1}{2}, t_r = 0$).*

*Proof.* The expected payoff of the two parties ($EU_r$ and $EU_l$) can be given as functions of $s_l(t_r, t_l)$. For interior values of $s_l(t_r, t_l)$, these payoffs are given by

$$EU_r(t_r, t_l) = [\frac{1}{s_l(t_r, t_l)} - 1](1 - t_r),$$

$$EU_l(t_r, t_l) = [2 - \frac{1}{s_l(t_r, t_l)}]t_l.$$

Here $s_l(t_r, t_l)$ is given by

$$s_l(t_r, t_l) = \frac{1}{2}(\frac{4\delta + \frac{2^{2\beta+2}}{4\beta+2}[t_l(1-t_l) - t_r(1-t_r)]}{(1-t_r)^2 - (1-t_l)^2})^{\frac{1}{2\beta}}.$$

I want to show that the payoff of the right party is decreasing in $t_r$ regardless of the strategy of the left party, which means that I want to show that $\frac{\partial EU_r(t_r,t_l)}{\partial t_r} < 0 \forall t_l$.[35] Conditional on $t_r = 0$ I then want to show that $\frac{\partial EU_l(t_r=0,t_l)}{\partial t_l} > 0$ for all feasible values of $t_l$, such that the only equilibrium is given by $t_r = 0$ and $t_l = t_{max} = \frac{1}{2}$. I will use two different approaches. For given parameter values $\beta$ and $\delta$ I will show analytically that $t_r = 0$ and $t_l = \frac{1}{2}$ is an equilibrium in this subgame, while I will use a simulation-based approach to show that Lemma 2 is satisfied for *all* feasible parameter values.

**Analytical solution**

I compute the derivative of the payoff of the right party with respect to $t_r$, which is given by

$$\frac{\partial EU_r(t_r, t_l)}{\partial t_r} = -[\frac{1}{s_l(t_r, t_l)} - 1] + (1-t_r)[\frac{-1}{s_l(t_r, t_l)^2}\frac{\partial s_l(t_r, t_l)}{\partial t_r}].$$

$\frac{\partial s_l(t_r,t_l)}{\partial t_r} > 0$ is a sufficient condition for $\frac{\partial EU_r(t_r,t_l)}{\partial t_r} < 0$. The derivative of $s_l(t_r, t_l)$ has the same sign as the derivative of $\chi(t_l, t_r) = [2s_l(t_r, t_l)]^{2\beta}$, and hence I compute the derivative of this easier expression given that $\beta \geq 1$ in this model, which is given by

[35]The probabilities above are only defined if $\frac{1}{s_l(t_r,t_l)} - 1 \in [0, 1]$. Otherwise the above probabilities are replaced with 0 or 1.

$$\frac{\partial \chi(t_l,t_r)}{\partial t_r} = \frac{2}{(t_l+t_r-2)^2}[\frac{-4\delta(t_r-1)}{(t_l-t_r)^2} - \frac{4^\beta}{2\beta+1}].$$

$\frac{\partial \chi(t_l,t_r)}{\partial t_r} > 0$ when

$$\frac{1-t_r}{(t_l-t_r)^2} > \frac{4^\beta}{(2\beta+1)}\frac{1}{4\delta}.$$

When $t_l$ and $t_r$ are close this is always satisfied. The smallest value of the left hand side is 4, which occurs when $t_l = \frac{1}{2}$ and $t_r = 0$. This implies that a sufficient condition for $\frac{\partial EU_r(t_r,t_l)}{\partial t_r}$ to be negative is given by

$$4 \geq \frac{4^\beta}{(2\beta+1)}\frac{1}{4\delta}.$$

Throughout the paper I construct examples where $\beta = 2$ and $\delta = \frac{1}{4}$, and the above inequality is satisfied for these parameter values.

**Best response for the left party** Given that $t_r = 0$ is the optimal strategy for the right party, I compute the best response for the left party. I will show that the expected payoff for the left party is increasing in $t_l$ given $t_r = 0$. I provide an analytical solution when $\beta = 2$ and $\delta = \frac{1}{4}$. The solution for other feasible parameter values follows the same pattern. The expected payoff of the left party is given by $EU_l(t_r = 0, t_l)$, which has a derivative given by

$$\frac{\partial EU_l(t_r=0,t_l)}{\partial t_l} = \frac{t_l[\frac{\frac{32(1-2t_l)}{5}}{1-(1-t_l)^2} - 2(1-t_l)\frac{(\frac{32t_l(1-t_l)}{5}+1)}{[1-(1-t_l)^2]^2}]}{2[\frac{\frac{32t_l(1-t_l)}{5}+1}{1-(1-t_l)^2}]^{\frac{5}{4}}} - \frac{2}{[\frac{\frac{32t_l(1-t_l)}{5}+1}{1-(1-t_l)^2}]^{\frac{1}{4}}} + 2.$$

This derivative is positive for any feasible $t_l$, which implies that the best response to $t_r = 0$ is given by $t_l = t_{max} = \frac{1}{2}$.

**Simulation-based solution**

I can also use a simulation-based approach to show that the lemma is satisfied for all feasible parameter values. I first want to show that $t_r = 0$ is the best response to any feasible $t_l$ given that $\beta$ and $\delta$ are drawn from their feasible regions.

1. I draw a random realization of $\beta^{rand}$, $\delta^{rand}$, $t_r^{rand}$ and $t_l^{rand}$ from their feasible regions.
2. Then I compare $EU_r(\beta^{rand}, \delta^{rand}, t_r^{rand}, t_l^{rand})$ and $EU_r(\beta^{rand}, \delta^{rand}, t_r = 0, t_l^{rand})$.
3. I repeat the process a large number of times.
4. I count how often $EU_r(\beta^{rand}, \delta^{rand}, t_r^{rand}, t_l^{rand}) > EU_r(\beta^{rand}, \delta^{rand}, t_r = 0, t_l^{rand})$.

If I can make arbitrarily many simulations without finding one random realization satisfying $EU_r(\beta^{rand}, \delta^{rand}, t_r^{rand}, t_l^{rand}) > EU_r(\beta^{rand}, \delta^{rand}, t_r = 0, t_l^{rand})$, I conclude that $t_r = 0$ is the best response to any feasible $t_l$. Then I want to show that $t_l = \frac{1}{2}$ is the best response to $t_r = 0$.

1. I draw a random realization of $\beta^{rand}$, $\delta^{rand}$ and $t_l^{rand}$ from their feasible regions.
2. Then I compare $EU_l(\beta^{rand}, \delta^{rand}, t_r = 0, t_l^{rand})$ and $EU_l(\beta^{rand}, \delta^{rand}, t_r = 0, t_l = \frac{1}{2})$.

3. I repeat the process a large number of times and count how many times $EU_l(\beta^{rand}, \delta^{rand}, t_r = 0, t_l^{rand}) > EU_l(\beta^{rand}, \delta^{rand}, t_r = 0, t_l = \frac{1}{2})$.

Again I conclude that $t_l = \frac{1}{2}$ is the best response to $t_r = 0$ if I cannot find one realization satisfying $EU_l(\beta^{rand}, \delta^{rand}, t_r = 0, t_l^{rand}) > EU_l(\beta^{rand}, \delta^{rand}, t_r = 0, t_l = \frac{1}{2})$.

□

**Lemma 3.** *The unique equilibrium in this subgame is for the right party to propose* $t_r = \tilde{t}(\beta, \delta) > 0$ *and win with probability* 1.

*Proof.* The two parties want to maximize

$$EU_r(t_r, t_l) = [\frac{1 - 2s_l(t_r, t_l)}{1 - s_l(t_r, t_l)}](1 - t_r),$$

$$EU_l(t_r, t_l) = [\frac{s_l(t_r, t_l)}{1 - s_l(t_r, t_l)}]t_l.$$

The expression for $s_l(t_r, t_l)$ is now given by

$$s_l(t_r, t_l) = \frac{1}{2}(\frac{-4\delta + \frac{2^{2\beta+2}}{4\beta+2}[t_l(1 - t_l) - t_r(1 - t_r)]}{(1 - t_r)^2 - (1 - t_l)^2})^{\frac{1}{2\beta}}.$$

The poorest voter maximally gets a payoff of $\frac{1}{4}\frac{2^{2\beta}}{4\beta+2} - \delta$ from voting for party $l$. The right party can hence set $t_r = \tilde{t}(\beta, \delta)$, implicitly defined by $\tilde{t}(\beta, \delta)[1 - \tilde{t}(\beta, \delta)]\frac{2^{2\beta}}{4\beta+2} = \frac{1}{4}\frac{2^{2\beta}}{4\beta+2} - \delta$ to win the election with certainty. Solving this equality leads to

$$\tilde{t}(\beta, \delta) = \frac{1}{2}[1 - 2\sqrt{\frac{(4\beta + 2)\delta}{2^{2\beta}}}].$$

I want to show that $t_r = \tilde{t}(\beta, \delta)$ and $t_l = \frac{1}{2}$ is the unique equilibrium in this subgame. By assumption I know that $t_l = \frac{1}{2}$ is a best response to $t_r = \tilde{t}(\beta, \delta)$ (because I have assumed that an indifferent party proposes its preferred tax rate), so it suffices to show that $t_r = \tilde{t}(\beta, \delta)$ is the best response to $t_l = \frac{1}{2}$ in order to establish the proposed solution as an equilibrium.

To show uniqueness, I show that there cannot exist an equilibrium where $t_l < \frac{1}{2}$ if $P_l(t_r, t_l) = 0$, because the left party proposes $t_l = \frac{1}{2}$ conditional on $P_l(t_r, t_l) = 0$. This means that when I look for equilibria involving $t_l < \frac{1}{2}$ I can restrict the attention to cases where $P_l(t_r, t_l) > 0$.

**Analytical proof**

I will analytically show that $t_r = \tilde{t}(\beta, \delta)$ and $t_l = \frac{1}{2}$ is an equilibrium conditional on $\beta = 2$ and $\delta = \frac{1}{4}$, but the proof will follow a similar pattern for other feasible parameter values. For these parameter values $\tilde{t}(\beta = 2, \delta = \frac{1}{4}) \approx 0.105$, so I need to show that the derivative is positive for all $t_r \leq \tilde{t}(\beta = 2, \delta = \frac{1}{4})$.

Inserting the parameter values into the expression for $s_l(t_r, t_l = \frac{1}{2})$ leads to

$$s_l(t_r, t_l = \frac{1}{2}) = \frac{1}{2}[\frac{-1 + \frac{32}{5}[0.25 - t_r(1 - t_r)}{((1 - t_r)^2 - 0.25)}]^{0.25}.$$

Then I can compute the derivative of $EU_r(t_r, t_l = \frac{1}{2})$. This derivative is messy, and is given on the following lines.

$$\frac{\partial EU_r(t_r, t_l = 0.5)}{\partial t_r} = -\frac{0.25(1 - \text{t\_r})\left(\frac{2(1-\text{t\_r})\left(\frac{32}{5}(0.25 - (1-\text{t\_r})\text{t\_r}) - 1\right)}{((1-\text{t\_r})^2 - 0.25)^2} + \frac{32(2\text{t\_r} - 1)}{5((1-\text{t\_r})^2 - 0.25)}\right)}{\left(\frac{\frac{32}{5}(0.25 - (1-\text{t\_r})\text{t\_r}) - 1}{(1-\text{t\_r})^2 - 0.25}\right)^{0.75}\left(1 - 0.5\left(\frac{\frac{32}{5}(0.25 - (1-\text{t\_r})\text{t\_r}) - 1}{(1-\text{t\_r})^2 - 0.25}\right)^{0.25}\right)}$$

$$+\frac{0.125(1-\text{t\_r})\left(1-1.\left(\frac{\frac{32}{5}(0.25-(1-\text{t\_r})\text{t\_r})-1}{(1-\text{t\_r})^2-0.25}\right)^{0.25}\right)\left(\frac{2(1-\text{t\_r})\left(\frac{32}{5}(0.25-(1-\text{t\_r})\text{t\_r})-1\right)}{((1-\text{t\_r})^2-0.25)^2}+\frac{32(2\text{t\_r}-1)}{5((1-\text{t\_r})^2-0.25)}\right)}{\left(\frac{\frac{32}{5}(0.25-(1-\text{t\_r})\text{t\_r})-1}{(1-\text{t\_r})^2-0.25}\right)^{0.75}\left(1-0.5\left(\frac{\frac{32}{5}(0.25-(1-\text{t\_r})\text{t\_r})-1}{(1-\text{t\_r})^2-0.25}\right)^{0.25}\right)^2}$$

$$-\frac{1-1.\left(\frac{\frac{32}{5}(0.25-(1-\text{t\_r})\text{t\_r})-1}{(1-\text{t\_r})^2-0.25}\right)^{0.25}}{1-0.5\left(\frac{\frac{32}{5}(0.25-(1-\text{t\_r})\text{t\_r})-1}{(1-\text{t\_r})^2-0.25}\right)^{0.25}}$$

This derivative is increasing when $t_r < \tilde{t}(\beta, \delta)$, which implies that the best response to $t_l = \frac{1}{2}$ is given by $t_r = \tilde{t}(\beta, \delta)$. Hence the proposed solution exist as an equilibrium.

**Simulation-based solution**

1. I first draw a random realization of $\beta^{rand}$, $\delta^{rand}$, $t_l^{rand}$ and $t_r^{rand}$ from their feasible regions.
2. Then I want to understand if $t_l^{rand}$ and $t_r^{rand}$ can be an equilibrium. There are two cases to consider.
   - If $P_l(\beta^{rand}, \delta^{rand}, t_l^{rand}, t_r^{rand}) > 0$ I check whether $EU_r(\beta^{rand}, \delta^{rand}, t_r^{rand}, t_l^{rand}) > EU_r(\beta^{rand}, \delta^{rand}, \tilde{t}(\beta, \delta), t_l^{rand})$, which means that I check if the right party can prefer another tax rate over $\tilde{t}(\beta, \delta)$.
   - If $P_l(\beta^{rand}, \delta^{rand}, t_l^{rand}, t_r^{rand}) = 0$, I know that the left party deviates to $t_l = \frac{1}{2}$. Given this deviation I draw random values of $t_r^{rand}$ and observe if $EU_r(\beta^{rand}, \delta^{rand}, t_r^{rand}, t_l = \frac{1}{2}) > EU_r(\beta^{rand}, \delta^{rand}, \tilde{t}(\beta, \delta), t_l = \frac{1}{2})$.
3. This process is repeated a large number of times.

The intuition is the same as in the simulation-based proof for Lemma 2. If an arbitrarily large number of simulations are drawn without finding one realization such that $t_r = \tilde{t}(\beta, \delta)$ is not the best response to $t_l$, I conclude that $t_r = \tilde{t}(\beta, \delta)$ is the best response to any feasible $t_l$. The best response to $t_r = \tilde{t}(\beta, \delta)$ is $t_l = \frac{1}{2}$, which means that the equilibrium is given by $t_l = \frac{1}{2}$ and $t_r = \tilde{t}(\beta, \delta)$.

□

**Proposition 1.** *The unique subgame perfect equilibrium is given by* $\{t_r^\star(c_r^\star, c_l^\star) = 0, c_r^\star = 1\}$ *and* $\{t_l^\star(c_r^\star, c_l^\star) = \frac{1}{2}, c_l^\star = 0\}$ *if* $P_r(\beta, \delta) \geq \frac{1-t^m(\beta)}{2}$. *Otherwise, the unique subgame perfect equilibrium is convergence to the median voter's preferences given by* $\{t_r^\star(c_r^\star, c_l^\star) = t^m(\beta), c_r^\star = 0\}$ *and* $\{t_l^\star(c_r^\star, c_l^\star) = t^m(\beta), c_l^\star = 0\}$.

*Proof.* In the subgames where $c_l = c_r$, both parties will propose $t^m(\beta)$, which leads to a payoff of $\frac{1-t^m(\beta)}{2}$ for the right party and $\frac{t^m(\beta)}{2}$ for the left party. If $c_r < c_l$ the right party proposes $\tilde{t}(\beta, \delta)$ and gets a payoff of $1 - \tilde{t}(\beta, \delta)$, while the left party gets 0. If $c_r > c_l$, the right party wins with probability $P_r(\beta, \delta) = \frac{2}{[\frac{4}{3}(\frac{2^{2\beta}}{4\beta+2}+4\delta)]^{\frac{1}{2\beta}}} - 1$ and the parties propose $t_r = 0$ and $t_l = \frac{1}{2}$.

I split the proof in two parts. First I show that $c_l = 1$ is dominated by $c_l = 0$ for the left party, and then I show that the right party will propose $c_r = 1$ if $P_r(\beta, \delta) \geq \frac{1-t^m(\beta)}{2}$. From Lemmas 1, 2 and 3 I know the outcomes in the four different subgames in the second stage of the game.

I here show that $c_l = 0$ is the best response to $c_r = 0$ and to $c_r = 1$. If $c_r = 1$, Lemma 1 shows that the left party get $\frac{t^m(\beta)}{2}$ by playing $c_l = 1$, and Lemma 2 shows that the left party get a payoff of $\frac{1-P_r(\beta,\delta)}{2}$ by playing $c_l = 0$. $P_r(\beta, \delta)$ is maximized when $\delta = \frac{1}{6}$. This means that the left party will never play $c_l = 1$ as a best response to $c_r = 1$ if the function $\xi(\beta) = 1 - P_r(\beta, \delta = \frac{1}{6}) - t^m(\beta)$ is positive for all values of $\beta$. I compute $\frac{\partial \xi(\beta)}{\partial \beta}$ and find that the minimum value of $\xi(\beta)$ occurs when $\beta = 2.37$. I verify that $\xi(\beta = 2.37) > 0$ and hence the left party will not play $c_l = 1$ as a best response to $c_r = 1$.

If $c_r = 0$, Lemma 3 shows that the left party get a payoff of $0$ from $c_l = 1$. Lemma 1 shows that the left party get a payoff of $\frac{t^m(\beta)}{2}$ by playing $c_l = 0$. Hence the left party will not play $c_l = 1$ as a best response to $c_r = 0$.

This implies that the equilibrium is given by $c_r = 1$ and $c_l = 0$ if the right party prefers $c_r = 1$ over $c_r = 0$, which is satisfied if $P_r(\beta, \delta) \geq \frac{1-t^m(\beta)}{2}$.

**Alternative proof** The above proof relies on Lemmas 2 and 3, which were proved analytically when $\beta = 2$ and $\delta = \frac{1}{4}$ and numerically for other parameter values. Here I will provide an analytical proof for Proposition 1 for certain parameter values without relying on the insights from these Lemmas. In particular, I prove Proposition 1 conditional on $\beta = 1$ and $\delta = \frac{1}{4}$.

**Equilibrium when $\beta = 1$ and $\delta = \frac{1}{4}$**

Suppose the subgame where $c_r > c_l$ is reached. The right party wins the election with probability $[\frac{1}{s_l(t_r,t_l)} - 1]$, where $s_l(t_r, t_l) = \frac{1}{2}(\frac{1+\frac{2^4}{6}[t_l(1-t_l)-t_r(1-t_r)]}{(1-t_r)^2-(1-t_l)^2})^{\frac{1}{2}}$. This expression is maximized when $t_l = \frac{1}{2}$ and $t_r = 0$, which leads to a maximum winning probability for the right party given by $0.34$. An upper bound for the expected payoff for the right party when $c_r > c_l$ is hence given by $0.34$. The left party wins the election with a probability not smaller than $0.66$ in this subgame, which means that the expected utility is bounded below by $0.66t^\star(\beta = 1)$.

Suppose the subgame where $c_l > c_r$ is reached. The poorest voter prefers $t_r = 0$ and $c_r = 0$ over $t_l = \frac{1}{2}$ and $c_l = 1$. I show this by inserting the parameter values into equation (1.3) for the poorest voter ($\theta_i = 0$), which leads to $U(t_r = 0, c_r = 0, \theta_i = 0, \beta = 1) = 0$ and $U(t_l = \frac{1}{2}, c_l = 1, \theta_i = 0, \beta = 1) = \frac{1}{6} - \frac{1}{4}$. This implies that the left party gets a payoff of $0$ if $c_l > c_r$.

Suppose one of the subgames where $c_r = c_l$ is reached. In these subgames convergence to the median voter's preferred tax rate of $t^{\star}(\beta = 1) = \frac{1}{5}$ is the only equilibrium.

Combining the outcome in the different subgames I show that $c_l = 1$ is a dominated strategy for the left party. Proposing $c_l = 1$ leads to a payoff of $\frac{t^{\star}(\beta=1)}{2}$ if $c_r = 1$ and 0 if $c_r = 0$. Proposing $c_l = 0$ leads to a payoff of at least $0.66t^{\star}(\beta = 1)$ if $c_r = 1$ and $\frac{t^{\star}(\beta=1)}{2}$ if $c_r = 0$. The right party gets a payoff of at least $\frac{1-t^{\star}(\beta=1)}{2} = 0.4$ by converging to the median voter's preferences, which is larger than the payoff in the diverging equilibrium. Hence the median solution is the only equilibrium.

$\square$

**Lemma 4.** *An increase in inequality increases the incentives to propose $c_r = 1$ if $\frac{\partial y(\beta,\delta)}{\partial \beta} = \frac{\partial P_r(\beta,\delta)}{\partial \beta} + \frac{\partial t^m(\beta)}{2\partial \beta} \geq 0$, where:*
*i) $\frac{\partial t^m(\beta)}{\partial \beta} > 0$,*
*ii) $\frac{\partial P_r(\beta,\delta)}{\partial \beta} \geq 0$ if $\beta$ is smaller than some threshold.*

*Proof.* The first derivative is given by

$$\frac{\partial t^m(\beta)}{\partial \beta} = 2^{2\beta+1} \frac{[\ln(2)(1 + 2\beta) - 1]}{[-2\beta + 2^{2\beta+1} - 1]^2}.$$

Evaluated in $\beta = 1$ (which is the lowest level of inequality in this model), the second derivative is given by

$$\frac{\partial P_r(\beta,\delta)}{\partial \beta}\Big|_{\beta=1} = \frac{\sqrt{3}}{\sqrt{4\delta + \frac{2}{3}}}[\frac{\ln(4\delta + \frac{2}{3})}{2} - \frac{\frac{4\ln(2)}{3} - \frac{4}{9}}{8\delta + \frac{4}{3}} - \frac{\ln(3)}{2} + \ln(2)].$$

This derivative is positive for all feasible values of $\delta$ evaluated in $\beta = 1$. This shows that the derivative is always positive when $\beta$ is below some threshold. I compute the limit when $\beta \to \infty$, and this limit is given by $\lim_{\beta\to\infty} P_r(\beta, \delta) = 0$. The function is continuous in $\beta$, and hence the derivative must be decreasing for larger values of $\beta$ (as long as $P_r(\beta, \delta)$ is positive for some intermediate values).

$\square$

**Proposition 2.** *The expected tax rate in the polarized equilibrium is lower than the preferred tax rate of the median voter for all values of $\delta$ if $\beta > \frac{3}{2}$. If $\beta \leq \frac{3}{2}$, the tax rate in the polarized equilibrium is lower than the preferred tax rate of the median voter as long as $\delta$ is small.*

*Proof.* The right party wants to propose a populist cultural policy if $P_r(\beta, \delta) \geq \frac{1-t^m(\beta)}{2}$. The tax rate is lower in a polarized equilibrium if

$$\mathbb{E}[t_{c_r=1}(\beta, \delta)] = [1 - P_r(\beta, \delta)]\frac{1}{2} < t^m(\beta). \tag{6}$$

I want to check if (6) is satisfied when $P_r(\beta, \delta) \geq \frac{1-t^m(\beta)}{2}$. A smaller $P_r(\beta, \delta)$ makes it harder to satisfy (6), so I will check if the inequality is satisfied when $P_r(\beta, \delta) = \frac{1-t^m(\beta)}{2}$. As long as $t^m(\beta) > \frac{1}{3}$, which happens if $\beta > \frac{3}{2}$, the expected tax rate is lower in a polarized equilibrium.

$\square$

## .2 Proofs for Section 1.4 (Extensions)

**Proposition 3.** *The equilibrium is given by a coalition between the populist party and the right party proposing $c_p = 1$ and $t_r = 0$ if $P_r(\beta, \delta) \geq \frac{1-t^m(\beta)}{2}$. Otherwise, the left and right parties form a coalition or rule in a minority government.*

*Proof.* Section 1.3.4 shows that party $l$ will never accept the coalition proposal.

If party $r$ accepts the proposal, the subgame from Section 1.3.4 is reached, and Lemma 2 shows that the unique equilibrium in this subgame is given by $t_r = 0$ and $t_l = \frac{1}{2}$. The right party prefers this solution over the median solution as long as $P_r(\beta, \delta) > \frac{1-t^m(\beta)}{2}$. $\square$

# .3 Proofs for Section 1.5 (Robustness)

**Proposition 4.** *The unique subgame perfect equilibrium is given by $t_r^\star = 0$, $c_r^\star = 1$ and $c_l^\star = 0$ if the right party gets the possibility to choose a cultural policy and $P_r(\beta, \delta) > 1 - 2t^m(\beta)$. Otherwise, the unique subgame perfect equilibrium is given by convergence to $t_l^\star = t_r^\star = t^m(\beta)$ and $c_l^\star = c_r^\star = 0$.*

*Proof.* In this version of the model there are also four subgames in the second stage of the model, depending on the choices of $c_r$ and $c_l$. I first show that the outcome within each subgame is the same as in the main model.

If $c_r = c_l$, then $t_r^\star = t_l^\star = t^m(\beta)$, which leads to payoffs $EU_r = 1 - t^m(\beta)$ and $EU_l = t^m(\beta)$.[36]

I now analyze the subgame where $c_r > c_l$. In the main model, I showed that $t_r^\star = 0$ in this subgame by showing that the derivative of $P_r(t_r, t_l)(1 - t_r)$ is decreasing in $t_r$ for all values of $t_l$. Hence it follows that the derivative of $P_r(t_r, t_l)(1 - t_r) + [1 - P_r(t_r, t_l)](1 - t_l)$ is also decreasing in $t_r$ for all values of $t_l$. Given that $t_r^\star = 0$, the left party maximizes $[1 - P_r(t_r = 0, t_l)]t_l$, and the main model shows that this expression is maximized by setting $t_l^\star = \frac{1}{2}$.

In the subgame where $c_l > c_r$, the equilibrium is given by $t_r^\star = \tilde{t}(\beta, \delta)$ and $t_l^\star = \frac{1}{2}$.

[36] A party can deviate to another tax rate without affecting the payoff, but I made the assumption that an indifferent party chooses the tax rate that maximizes the winning probability.

I use a simulation-based approach to show that this is outcome for all parameter values.

I then analyze the first stage of the model. The left party will never propose $c_l = 1$ because $t^m(\beta) > \tilde{t}(\beta, \delta)$. The right party gets a payoff of $1 - t^m(\beta)$ when proposing $c_r = 0$, and a payoff of $P_r(\beta, \delta) + \frac{1-P_r(\beta,\delta)}{2} = \frac{1}{2} + \frac{P_r(\beta,\delta)}{2}$ when proposing $c_r = 1$. Hence, there is a unique subgame perfect equilibrium where $c_r > c_l$ if $P_r(\beta, \delta) > 1 - 2t^m(\beta)$. Otherwise, the equilibrium involves convergence.

□

**Proposition 5.** *There is an equilibrium where $c_l^\star = 1$ and $c_r^\star = 0$ if $\alpha_{max} \geq \frac{1}{2[1-t^m(\beta)]}$. In this equilibrium, the right party proposes $t_r^\star = \tilde{t}(\beta, \delta) < t^m(\beta)$, while the left party proposes $t_l^\star = \frac{1}{2}$.*

*Proof.* In the subgame where $c_l > c_r$ the right party wants to maximize $\frac{1}{\alpha_{max}} \frac{\frac{1}{2} - s_l(t_r,t_l)}{1-s_l(t_r,t_l)}[1 - t_r]$.

For interior probabilities, this maximization problem does not depend on the size of $\alpha_{max}$, and hence the right party proposes $t_r = \tilde{t}(\beta, \delta)$ and wins with probability $\frac{1}{2\alpha_{max}}$. The left party proposes $t_l = \frac{1}{2}$ and wins with probability $1 - \frac{1}{2\alpha_{max}}$.

I then analyze the first-stage choices of cultural policies. In the subgame where $c_l > c_r$ the right party gets a payoff of $\frac{1-\tilde{t}(\beta,\delta)}{2\alpha_{max}}$, while the left party gets $\frac{1}{2} - \frac{1}{4\alpha_{max}}$. If one of the parties deviate a subgame where $c_r = c_l$ is reached. In these subgames the right party gets $\frac{1-t^m(\beta)}{2}$ and the left party gets $\frac{t^m(\beta)}{2}$. The payoff of the right party is always larger in the subgame where $c_r < c_l$ than in the subgames where $c_r = c_l$. The payoff of the left party is larger if $\frac{1}{2} - \frac{1}{4\alpha_{max}} > \frac{t^m(\beta)}{2}$, which can be simplified to $\alpha_{max} > \frac{1}{2(1-t^m(\beta))}$.

□

## .4 Proofs for Section 1.6 (A model with exogenous tax preferences)

**Proposition 6.** *Suppose a majority of voters prefer $t_{high}$. The unique subgame perfect equilibrium is given by $c_r^\star = 1$, $c_l^\star = 0$, $t_r^\star = t_{low}$ and $t_l^\star = t_{high}$ if $\frac{U_p}{U} \geq max[\frac{1}{1-P_{1,0}}, \frac{1}{2P_{1,0}}]$. Otherwise, the only pure subgame perfect equilibrium is given by $c_r^\star = c_l^\star = 0$ and $t_l^\star = t_r^\star = t_{high}$.*

*Proof.* There are $3$ different subgames to consider.

1. $c_r = c_l$. In these subgames, both parties propose $t = t_{high}$ as $t_{low}$ leads to a payoff of $0$.

2. $c_r > c_l$. Suppose there is an equilibrium where $t_r = t_{low}, t_l = t_{high}$. The right party gets $P_{1,0}U_p$, while the left party gets $(1 - P_{1,0})U$. The right party gets a payoff of $0$ from deviating to $t_{high}$, while the left party gets $U$ from deviating to $t_{low}$. Hence $(1 - P_{1,0})U_p > U$ is necessary for divergence in tax rates to be an equilibrium in this subgame.

3. $c_l > c_r$ If $P_{0,1}U_p > U$, there is an equilibrium with divergence in this subgame, where the left party gets a payoff of $[1 - P_{0,1}]U_p$. Otherwise, there is only a mixed equilibrium where the left party gets a payoff *smaller* than $[1 - P_{0,1}]U_p$.

   **First stage solution** I first show that proposing $c_l = 1$ is a dominated strategy for the left party as long as $(1 - P_{1,0})U_p > U$. If $c_r = 0$, the left party gets maximum $[1 - P_{0,1}]U_p$ from $c_l = 1$, which is smaller that the payoff of $\frac{U_p}{2}$ from $c_l = 0$. If $c_r = 1$, the left party gets $\frac{U_p}{2}$ from $c_l = 1$, which is smaller than the maximum payoff of $[1 - P_{1,0}]U_p$ from playing $c_l = 0$.

The right party prefers $c_r = 1$ over $c_r = 0$ if $P_{1,0}U_p \geq \frac{U}{2}$.

If $P_{1,0}U_p < \frac{U}{2}$ and $(1 - P_{1,0})U_p > U$, there is a pure equilibrium given by complete convergence.

I now analyze if there can exist other pure equilibria. In the subgame where $c_l > c_r$, the only potential pure equilibrim is given by $t_r = t_{low}$ and $t_l = t_{high}$. In the first stage, the left party gets a payoff of $1 - P_{0,1}U_p < \frac{U_p}{2}$ when $c_r < c_l$, and can deviate to $c_l = 1$ and get $\frac{U_p}{2}$. Hence there cannot be a pure subgame perfect equilibrium involving the subgame where $c_l > c_r$. In the subgame where $c_r > c_l$, the only potential pure equilibrium is given by $t_r = t_{low}$ and $t_l = t_{high}$. In the subgames where $c_r = c_l$, the only pure equilibrium is given by $t_r = t_l = t_{high}$.

□

**Proposition 7.** *Suppose the median voter is close to indifferent between $t_{high}$ and $t_{low}$. $P_{1,0} \geq 1 - P_{0,1}$ as long as $G[\cdot]$ is concave.*

*Proof.* I analyze how many voters that prefer $c_r = 1$, $t_r = t_{low}$ over $c_l = 0$ and $t_l = t_{high}$. I first analyze the group of size $\alpha$ that prefers $c = 1$, and voters belonging to this group vote for the right party as long as

$$w_i(1 - t_{low}) + T(t_{low}) + \delta > w_i(1 - t_{high}) + T(t_{high}).$$

This inequality is satisfied for a share $1 - G[\frac{T(t_{high}) - T(t_{low}) - \delta}{t_{high} - t_{low}}]$ of these voters.

I then consider the other group (of size $1 - \alpha$). Voters belonging to this group prefer the right party if

$$w_i(1 - t_{low}) + T(t_{low}) - \delta > w_i(1 - t_{high}) + T(t_{high}).$$

This leads to a vote share for the right party given by $s_{1,0} = \alpha(1 - G[\frac{T(t_{high}) - T(t_{low}) - \delta}{t_{high} - t_{low}}]) + (1 - \alpha)(1 - G[\frac{T(t_{high}) - T(t_{low}) + \delta}{t_{high} - t_{low}}])$ when the right party proposes $t_r = t_{low}$, $c_r = 1$ and the left party proposes $t_l = t_{high}$, $c_l = 0$.

I then find the probability that $s_{1,0} > \frac{1}{2}$. This occurs if $\alpha(1 - G[\frac{T(t_{high}) - T(t_{low}) + \delta}{t_{high} - t_{low}}]) + (1 - \alpha)(1 - G[\frac{T(t_{high}) - T(t_{low}) - \delta}{t_{high} - t_{low}}]) < \frac{1}{2}$, which happens if

$$\alpha > \underline{\alpha^{right}} = \frac{G[\frac{T(t_{high}) - T(t_{low}) + \delta}{t_{high} - t_{low}}] - \frac{1}{2}}{G[\frac{T(t_{high}) - T(t_{low}) + \delta}{t_{high} - t_{low}}] - G[\frac{T(t_{high}) - T(t_{low}) - \delta}{t_{high} - t_{low}}]}$$

.

I can repeat the same analysis to find the vote share of the right party when the left party proposes $t_l = t_{high}$, $c_l = 1$ and the right party proposes $t_r = t_{low}$, $c_r = 0$. This share is given by $s_{0,1} = \alpha(1 - G[\frac{T(t_{high}) - T(t_{low}) + \delta}{t_{high} - t_{low}}]) + (1 - \alpha)(1 - G[\frac{T(t_{high}) - T(t_{low}) - \delta}{t_{high} - t_{low}}])$.

The left party wins the election if $s_{0,1} < \frac{1}{2}$, which happens if

$$\alpha > \underline{\alpha^{left}} = \frac{\frac{1}{2} - G[\frac{T(t_{high}) - T(t_{low}) - \delta}{t_{high} - t_{low}}]}{G[\frac{T(t_{high}) - T(t_{low}) + \delta}{t_{high} - t_{low}}] - G[\frac{T(t_{high}) - T(t_{low}) - \delta}{t_{high} - t_{low}}]}$$

.

I now show that $\underline{\alpha^{left}} > \underline{\alpha^{right}}$, which is equivalent with $\frac{1}{2} - G[\frac{T(t_{high}) - T(t_{low}) - \delta}{t_{high} - t_{low}}] > G[\frac{T(t_{high}) - T(t_{low}) + \delta}{t_{high} - t_{low}}] - \frac{1}{2}$ or

$$G[\frac{T(t_{high}) - T(t_{low}) + \delta}{t_{high} - t_{low}}] + G[\frac{T(t_{high}) - T(t_{low}) - \delta}{t_{high} - t_{low}}] < 1. \qquad (7)$$

I assumed that the median voter is close to indifferent, which means that $G[\frac{T(t_{high})-T(t_{low})}{t_{high}-t_{low}}] = \frac{1}{2}$. Equation (7) is then satisfied for all concave $G[\cdot]$, because all concave functions satisfy $\frac{f(x)}{2} + \frac{f(y)}{2} \leq f(\frac{x+y}{2})$.

□

## .5 Proofs for Section 1.7 (Appendix B. Diminishing marginal utility of income )

**Lemma 5.** *A marginal increase in $\rho$ from $\rho = 0$ will affect effort according to*

$$\frac{\partial e^{\star}(t, \theta_i, \rho)}{\partial \rho}\Big|_{\rho=0} = \frac{(1-t)\theta_i^{\beta}}{2}[\ln(2) - 2\ln[(1-t)\theta^{\beta}]].$$

*The effect on tax revenues is given by*

$$\frac{\partial T(t, \rho)}{\partial \rho}\Big|_{\rho=0} = T(t)[1 + 0.692(1-2\beta) - \frac{1}{2\beta+1} - 2\ln(1-t)]$$

*Proof.* The effort level and tax revenues can be computed as functions of $\rho$.

$$e^{\star}(t, \theta, \beta, \rho) = \frac{1}{2^{\frac{1}{1+\rho}}}[(1-t)\theta^{\beta}]^{\frac{1-\rho}{1+\rho}}$$

$$\frac{\partial e^{\star}(t, \theta, \beta, \rho)}{\partial \rho}\Big|_{\rho=0} = \frac{(1-t)\theta^{\beta}}{2}[\ln(2) - 2\ln[(1-t)\theta^{\beta}]]$$

This expression equals zero when $(1-t)\theta^{\beta} = \sqrt{2}$.

I also take the derivative of $T(t, \beta, \rho)$ and evaluate this derivative in $\rho = 0$.

$$T(t,\beta,\rho) = \frac{t}{2}\frac{1}{2^{\frac{1}{1+\rho}}}[(1-t)]^{\frac{1-\rho}{1+\rho}}\frac{1+\rho}{2\beta+1+\rho}2^{\frac{2\beta+1+\rho}{1+\rho}}$$

This derivative is given by

$$\frac{\partial T(t,\beta,\rho)}{\partial\rho}\Big|_{\rho=0} = T(t,\beta)[1+0.692(1-2\beta) - \frac{1}{2\beta+1} - 2\ln(1-t)]$$

□

# .6 Proofs for Section 2.2 (Identification and estimation of unbiased treatment effects)

**Proposition 8.** *The correlation between the change between time $t^\star$ and $t^\star + s$ and the change between $t^\star$ and $t^\star + \delta$, approaches zero when the outcome follows a Lévy process with continuous paths, and $s \to 0$ and $\delta > 0$.*

*Proof.* The correlation is given by

$$corr[f(t^\star{+}\delta){-}f(t^\star{-}s), f(t^\star{+}s){-}f(t^\star{-}s)] = corr[f(t^\star{+}\delta), f(t^\star{+}s)] = \sqrt{\frac{min[s,\delta]}{max[s,\delta]}}$$

We can observe that $\lim_{s\to 0}\sqrt{\frac{min[s,\delta]}{max[s,\delta]}} = 0$. □

## .7 Proofs for Section 2.5 (Causal inference with interference at the population level)

**Proposition 9.** *Assume that SUTVA is not satisfied. As long as assumption 2 is satisfied we can estimate an unbiased treatment effect if the treatment group is small, but we need to make more assumptions to estimate an unbiased population-level treatment effect.*

*Proof.* The first part of the proof follows directly by observing that $\lim_{\hat{k}\to 0}\alpha(t+s,\hat{k}) = \lim_{\hat{k}\to 0}[f_1[t+s,\hat{k}] - f_0[t+s,\hat{k}]] = f_1[t+s,0] - f_0[t+s,0] = \alpha(t+s,0)$. Assume that the second part of the proposition is false, which means that we do not need to make more assumptions than assumption 2 to identify an unbiased population-level treatment effect. This means that there exists a $\hat{k}$ such that $f_1[t+s,\hat{k}] - f_0[t+s,\hat{k}] = f_1[t+s,N] - f_0[t+s,0]$. Letting $\hat{k}\to N$ means that we will estimate $\lim_{\hat{k}\to N} f_1[t+s,\hat{k}] - \lim_{\hat{k}\to N} f_0[t+s,\hat{k}]$. We have assumed that the first term approaches $f_1(t+s,N)$, which means that we get an unbiased treatment effect if and only if $\lim_{\hat{k}\to N} f_0[t+s,\hat{k}] = f_0[t+s,0]$. This is not satisfied without making more assumptions, and hence we have reached a contradiction for $\hat{k}\to N$. The same argument applies for $\hat{k}\to 0$. An intermediate $\hat{k} = \hat{k}'$ will estimate $f_1[t+s,\hat{k}'] - f_0[t+s,\hat{k}']$. Analyzing the two terms separately leads to the same contradiction as above. Alternatively we can analyze both terms together, and this will lead to an unbiased treatment effect if and only if $f_1[t+s,\hat{k}'] - f_0[t+s,\hat{k}'] = f_1[t+s,N] - f_0[t+s,0]$. For this to be satisfied we need to make further assumptions about $f_1(\cdot)$ and $f_0(\cdot)$, and hence we have reached a contradiction. $\square$

**Proposition 10.** *The relative importance of the bias ($y[b(f_0),\alpha]$) is decreasing in $\alpha$, and it will be increasing in $f_0$ if $b(f_0)$ is a monotonic and differentiable function that satisfies $b(0) = 0$.*

*Proof.*

$$\frac{\partial y[b(f_0),\alpha]}{\partial \alpha} = \frac{-|b(f_0)|}{[\alpha + b(f_0)]^2} \leq 0$$

$$\frac{\partial y[b(f_0),\alpha]}{\partial f_0} = \frac{\alpha b(f_0)}{|b(f_0)|[\alpha + b(f_0)]^2} \frac{\partial b(f_0)}{\partial f_0} \geq 0$$

Any monotonic and differentiable function satisfies $b(f_0)\frac{\partial b(f_0)}{\partial f_0} \geq 0$ as long as $b(0) = 0$.

□

# .8 Proofs for Section 3.3 (Analysis of a basic model)

**Proposition 11.** *Let $q^\star$ be the solution to $q^\star = F[bp^v(q^\star)]$, where $F(\tilde{c}) = q^\star$. A Bayesian Nash Equilibrium is given by*

$$r(c_i) = \begin{cases} 1 & \text{if } c_i \leq \tilde{c}, \\ 0 & \text{if } c_i > \tilde{c}. \end{cases}$$

*Proof.* Suppose there is an equilibrium where all victims report if and only if $c_i \leq \tilde{c}$. If this is an equilibrium, the share of victims reporting is given by $q^\star = F(\tilde{c})$, which leads to expected benefits of $bp^v(q^\star)$ from reporting. All victims with $c_i \leq \tilde{c}$ get an expected payoff of $bp^v(q^\star) - c_i \geq 0$ from choosing $r(c_i) = 1$, and will not deviate to $r(c_i) = 0$ and get a payoff of $0$. All victims with $c_i > \tilde{c}$ get a payoff of $bp^v(q^\star) - c_i < 0$ from $r(c_i) = 1$, which is smaller than the payoff of $0$ from $r(c_i) = 0$. Hence, the proposed solution is an equilibrium.

The function $p^v(q)$ is continuous in $q$, and hence $F[p^v(q)]$ is a continuous mapping from $[0,1]$ to $[0,1]$. By Brouwer's fixed point theorem there is a $q^\star$ such that $q^\star = F[bp^v(q^\star)]$.

$\square$

**Proposition 12.** *Suppose there is an equilibrium where $q^\star > 0$. Consider a reporting strategy given by*

$$r^\star(c_i) = \begin{cases} 1 & \textit{if } c_i \leq \tilde{c} + \Delta, \\ 0 & \textit{if } c_i > \tilde{c} + \Delta. \end{cases}$$

*For any small $\Delta$, $W_{r^\star(c_i)} \geq W_{r(c_i)}$.*

*Proof.* The social welfare in equilibrium is given by

$$W_{r(c_i)}(q^\star) = \Pi[\int\limits_{c_i=\underline{c}}^{c_i=F^{-1}(q^\star)} [bp^v(q^\star) - c_i]F'(c_i)dc_i] - \int\limits_{c_i=\underline{c}}^{c_i=\overline{c}} [s_w()C_w F'(c_i)dc_i].$$

The derivative of the social welfare function with respect to $q^\star$ is found using the Leibniz integral rule, which leads to

$$\frac{\partial W_{r(c_i)}(q^\star)}{\partial q^\star} = \Pi[\frac{\partial F^{-1}(q^\star)}{\partial q^\star}[bp^v(q^\star) - F^{-1}(q^\star)]F'[F^{-1}(q^\star)] + \int\limits_{c_i=\underline{c}}^{c_i=F^{-1}(q^\star)} b\frac{\partial p^v(q^\star)}{\partial q^\star}F'(c_i)dc_i].$$

Using the formula for the derivative of inverse function, the above expression simplifies to

$$\frac{\partial W_{r(c_i)}(q^\star)}{\partial q^\star} = \Pi[bp^v(q^\star) - F^{-1}(q^\star) + b\frac{\partial p^v(q^\star)}{\partial q^\star}F[F^{-1}(q^\star)]].$$

In equilibrium $bp^v(q^\star) = F^{-1}(q^\star)$, which leads to

$$\frac{\partial W_{r(c_i)}(q^\star)}{\partial q^\star} = \Pi b \frac{\partial p^v(q^\star)}{\partial q^\star} q^\star.$$

Hence, a small increase in reporting will increase social welfare.

I now construct a reporting strategy $r^\star(c_i)$ that equals $r(c_i)$ for all costs except that victims with costs between $\tilde{c}$ and $\tilde{c}+\Delta$ switch to $r(c_i) = 1$. This strategy implements a marginally larger $q^\star$ as long as $\Delta$ is small, and will increase social welfare.

$\square$

**Lemma 6.** *The probability of being wrongfully convicted is* $s_w(\tilde{k}) = 1 - \sum_{i=0}^{i=\tilde{k}-1} e^{-\gamma} \frac{\gamma^i}{i!}$, *and* $s_w(\tilde{k})$ *is decreasing in* $\tilde{k}$.

*Proof.* By modeling the population to consist of a continuum of agents, the Poisson Limit Theorem can be used to show that a binomial distribution approximates a Poisson distribution.

A false victim makes a report against each agent with probability $\frac{1}{n-1}$. In a population of $n$ agents, there are $\gamma n$ agents making false reports. Hence, the number of reports made against each agent follows a binomial distribution. This paper considers a large population with a continuum of agents. Let $P(k|t = I)$ denote the probability of observing $k$ reports against a non-criminal (type $t = I$). As $n \to \infty$, the Poisson Limit Theorem says that

$$P(k|t = I) = \lim_{n\to\infty} \binom{\gamma n}{k} \frac{1}{[n-1]^k} [1 - \frac{1}{n-1}]^{\gamma n - k} = \frac{e^{-\gamma}\gamma^k}{k!}.$$

Hence the probability of being wrongfully convicted when $\tilde{k}$ reports is necessary for conviction follows a Poisson distribution with parameter $\gamma$. As this paper considers a continuum of agents, I will use this approximation.

$\square$

**Proposition 13.** *The Information Escrow maximizes social welfare for any $\tilde{k}$.*

*Proof.* Social welfare is given by $W = \int_{\underline{c}}^{\overline{c}} [\Pi() u(c_i) - s_w() C_w] F'(c_i) dc_i$. The term $s_w() C_w$ depends on the behavior of exogenous false victims, and is hence constant for given values of $\tilde{k}$.

This means that social optimum is found by maximizing $\int_{\underline{c}}^{\overline{c}} \Pi() u(c_i) F'(c_i) dc_i$, which is the payoff for the realized victims of crime.

For each realized victim $i$, I now characterize the ideal outcome for different states of the world. There are two cases to consider.

1. If there are $k \geq \tilde{k}$ victims of the criminal, then victim $i$ gets benefits $b - c_i$ if $r(c_i) = 1$ and at least $\tilde{k} - 1$ other victims report. In this case the ideal outcome for $i$ is $r(c_i) = 1$ and at least $\tilde{k} - 1$ of the other victims also reporting.

2. If there are $k < \tilde{k}$ victims, then victim $i$ gets benefits $-c_i$ from $r(c_i) = 1$. In this case the ideal outcome is to set $r(c_i) = 0$.

The Information Escrow replicates the above characterization, and hence leads to the optimal outcome for each $i$ maximizes social welfare. $\square$

## .9 Proofs for Section 3.4 (Analysis of the general model)

**Proposition 14.** *A Bayesian jury observing $\hat{k}$ reports chooses to convict the defendant if and only if $\hat{k} \geq k^{\star}(q, q_f, \overline{p})$.*

*Proof.* The jury can compute the probability of observing $k$ reports against an innocent defendant. I let $t = G$ denote a criminal and $t = I$ denote an innocent agent. Using the procedure from Lemma 6, I know that $P(k|t = I)$ follows a Poisson distribution for all reporting strategies of false victims.

The posterior probability of guilt after observing $k$ reports is given by

$$P[t = G|k] = \frac{\pi P[k|t = G]}{\pi P[k|t = G] + (1 - \pi)P(k|t = I)}.$$

The jury chooses to convict the defendant if $P[t = G|k] > \overline{p}$.

I want to show that the posterior probability of guilt is increasing in $k$, which happens if

$$\frac{P(k + 1|t = G)}{P(k + 1|t = I)} > \frac{P(k|t = G}{P(k|t = I)}.$$

I rewrite this expression as

$$P(k + 1|t = G)P(k|t = I) > P(k|t = G)P(k + 1|t = I). \tag{8}$$

I now want to show that this inequality is satisfied. It is possible for false reports to be directed towards true criminals, which means that the behavior of false victims is included in $P(k|t = G)$. I first split the probability $P(k|t = G)$ into the behavior of true and false victims. Observing $k+1$ reports against a guilty offender can arise in different settings. Given a share $q$ and $q_f$ of true and false victims choosing to report, the probability of observing $k$ *truthful* reports against a criminal is given by $P_c(k,q)$, which, for simplicity, I write as $P_c(k)$. The probability of observing $k$ *false* reports against an guilty agent similarly depends on $q_f$, but to simplify the notation I write this expression as $P(k|t=I)$. This means that $P(k+1|t=G) = P_c(k+1)P(0|t=I) + P_c(k)P(1|t=I) + P_c(k-1)P(2|t=I)....P_c(0)P(k+1|t=I)$. I insert the above expressions into Equation (8) and get

$$P_c(k+1)P(0|t=I)P(k|t=I)+P_c(k)P(1|t=I)P(k|t=I)....P_c(0)P(k+1|t=I)P(k|t=I)$$

$$> P_c(k)P(0|t=I)P(k+1|t=I)+P_c(k-1)P(1|t=I)P(k+1|t=I)...P_c(0)P(k|t=I)P(k+1|t=I)$$

A sufficient condition for this inequality to be satisfied is given by

$$P_c(k)[P(1|t=I)P(k|t=I) - P(k=0|t=I)P(k+1|t=I)]$$

$$+P_c(k-1)[P(2|t=I)P(k|t=I) - P(1|t=I)P(k+1|t=I)]$$

$$... + P_c(0)[P(k+1|t=I)P(k|t=I) - P(k|t=I)P(k+1|t=I)] > 0$$

This inequality is satisfied if $P(j|t=I)P(k|t=I) > P(j-1|t=I)P(k+1) \quad \forall j \in \{1, 2..k+1\}$.

$P(j|t=I)$ follows a Poisson distribution, so the above inequality can be rewritten as

$$e^{-\gamma}\frac{\gamma^j}{j!}e^{-\gamma}\frac{\gamma^k}{k!} > e^{-\gamma}\frac{\gamma^{j-1}}{(j-1)!}e^{-\gamma}\frac{\gamma^{k+1}}{(k+1)!}$$

$$\gamma^{j+k}(\frac{1}{j} - \frac{1}{k+1}) > 0$$

This is always satisfied when $j < k+1$ and holds with equality when $j = k+1$, which means that the probability of guilt is increasing in $k$ for any given $q$.

Hence, the optimal strategy of the jury is to make a conviction if the number of reports, $k$, exceeds some threshold $k^\star(q, q_f, \overline{p})$. This threshold may be arbitrarily large.[37]

$\square$

**Proposition 15.** *In a Perfect Bayesian Equilibrium, a Bayesian jury convicts the defendant after observing at least $k^\star(q, q_f, \overline{p})$ reports. The reporting strategy of true victims is given by*

$$r_v(c_i) = \begin{cases} 1 & \text{if } c_i \leq \hat{c}, \\ 0 & \text{if } c_i > \hat{c}. \end{cases}$$

*The reporting strategy of false victims is given by*

$$r_f(c_i) = \begin{cases} 1 & \text{if } c_i \leq \hat{c}_f, \\ 0 & \text{if } c_i > \hat{c}_f. \end{cases}$$

*The thresholds $\hat{c}$ and $\hat{c}_f$ are given by $\hat{c} = F^{-1}(q^\star)$ and $\hat{c}_f = F^{-1}(q^\star_f)$, where $q^\star = F[bp^v(q^\star, q^\star_f)]$ and $q^\star_f = F[bp^f(q^\star, q^\star_f)]$.*

*Proof.* I first assume $p^v(q, q_f)$ is continuous around the thresholds when the share of victims reporting is given by $q^\star$ and $q^\star_f$, and use an approach that mimics the proof

[37] I only showed that the posterior probability is increasing in $k$, but did not show that there exists a $\hat{k}$ such that $P(t = G|\hat{k}) > \overline{p}$.

of Proposition 11.

Suppose there is an equilibrium where true victims report if $c_i \leq \hat{c}$ and false victims report if $c_i \leq \hat{c}_f$. If this is an equilibrium, then the conviction probabilities are given by $p^v(q^\star, q_f^\star)$ and $p^f(q^\star, q_f^\star)$ such that the expected benefits from reporting are given by $bp^v(q^\star, q_f^\star)$ and $bp^f(q^\star, q_f^\star)$. Given these expected benefits, all true victims with $c_i \leq \hat{c}$ get a payoff of $bp^v(q^\star, q_f^\star) - c_i \geq 0$ from reporting, while all true victims with $c_i > \hat{c}$ get a payoff of $bp^v(q^\star, q_f^\star) - c_i < 0$ from reporting. A similar argument holds for the false victims.[38]

The jury may observe $k = 0, 1, 2....$ reports. A Perfect Bayesian Equilibrium specifies beliefs in any of these information sets. Bayesian jury behavior is given by conviction if $k \geq k^\star(q, q_f, \overline{p})$. Given the strategies by true and false victims, the strategy following from Bayesian beliefs leads to conviction probabilities for true and false victims given by $p^v(q^\star, q_f^\star) = P^v[k \geq k^\star(q^\star, q_f^\star, \overline{p})|q^\star, q_f^\star]$ and $p^f(q^\star, q_f^\star) = P^f[k \geq k^\star(q^\star, q_f^\star, \overline{p})|q^\star, q_f^\star]$. Hence each belief is updated to the strategies of other players, and each strategy is optimal given these beliefs.

As long as $q > 0$ or $q_f > 0$, the jury can use Bayes' rule to compute the probability of guilt, but if $q = q_f = 0$, reporting is not observed in equilibrium. I made the assumption that a share $\epsilon$ of false victims exogenously choose to report, which means that the jury sets a conviction threshold of $\lim_{q \to 0} k^\star(q, q_f, \overline{p}) = \infty$ in this case. This implies that the threshold $\hat{c} = \hat{c}_f = \underline{c}$ is a Perfect Bayesian Equilibrium combined with the beliefs $q^\star = 0$, $q_f^\star = 0$. This outcome is always an equilibrium, and hence it is also an equilibrium if $p^v(q, q_f)$ is discontinuous. This equilibrium satisfies the requirements from Proposition 15. All victims report if $c_i \leq \underline{c}$, where $F(\underline{c}) = 0$. Given this behavior, the jury sets a conviction threshold of $\lim_{q \to 0} k^\star(q, q_f, \overline{p}) = \infty$,

[38]Given these expected benefits, all false victims with $c_i \leq \hat{c}_f$ get a payoff of $bp^f(q^\star, q_f^\star) - c_i \geq 0$ from reporting, while all false victims with $c_i > \hat{c}_f$ get a payoff of $bp^f(q^\star, q_f^\star) - c_i < 0$ from reporting.

which means that the strategy to report if and only if $c_i \leq \underline{c}$ is optimal for all victims.

$\square$

**Proposition 16.** *Suppose $t_r$ can be arbitrarily large. If $\pi < 1 - P(k = 1)$, the pledging mechanism will implement an outcome where $q \to 1$, $q_f \to 0$ and $p^v(r) \to 1$, which will maximize social welfare.*

*Proof.* Suppose there is a separating equilibrium where all true victims report, while no false victims report.

The true victim with the highest cost of reporting, given by $c_i = \overline{c}$, will not deviate as long as

$$p^v(r)[b + \frac{t_r}{p^v(r)}] - \overline{c} - t_r = p^v(r)b - \overline{c} \geq 0.$$

I have assumed that $b > \overline{c}$, which means that the above inequality is satisfied as long as $p^v(r)$ is sufficiently close to $1$.

The false victim with the lowest cost of reporting, given by $c_i = \underline{c}$, will not deviate from not reporting as long as

$$p^f(r)[b + \frac{t_r}{p^v(r)}] - \underline{c} - t_r < 0.$$

This inequality can be rewritten as

$$\frac{p^f(r) - p^v(r)}{p^v(r)} t_r + p^f(r)b - \underline{c} < 0. \tag{9}$$

$p^f(r) - p^v(r) < 0$ when $\pi < 1 - P(k = 1)$, which implies that Equation (9) is satisfied is satisfied for an arbitrarily large value of $t_r$. Given an arbitrarily large value of $t_r$, it is possible to sustain a separating equilibrium also when $p^v(r)$ is close to one.

The above argument shows that there is a separating equilibrium. Now I show that this separating equilibrium maximizes social welfare.

As described in previous sections, social welfare is given by

$$\int_{\underline{c}}^{\overline{c}} [\Pi()u(c_i) - s_w()C_w]F'(c_i)dc_i.$$

Using the pledging mechanism, $s_w() = 0$, which means that the pledging mechanism implements the social optimum if welfare is maximized for each realized victim.

By assumption, $b > \overline{c}$, which means that all victims want to report if a report leads to conviction.

This mechanism implements $p^v(r)$ close to 1. Each victim gets a payoff of $p^v(r)b - c_i$, which approximates the optimal outcome for each $i$ as $p^v(r) \to 1$.

$\square$