Essays in Public Finance

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Robert Scherf

aus Münster

2019

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Martin Hellwig
Zweitreferent:	Prof. Dr. Felix Bierbrauer
Tag der mündlichen Prüfung:	12. Juli 2019

Acknowledgements

My supervisors Martin Hellwig and Felix Bierbrauer have introduced me to the subject of public finance and taught me all I know about economic research. I am grateful for all their advice, guidance, and patience, as well as their support of my plans to spend the last three years of my doctorate in the United States.

Teaching has been one of the most rewarding experiences of my studies. I want to thank Barry Nalebuff for his mentorship, as well as Steve Berry and Dezsö Szalay for giving me the opportunity to teach in their courses. I am also grateful to Matthew Weinzierl, who not only inspired the first chapter of this dissertation but also hired me as his Research Associate during the end of my doctorate. Matt has been a mentor both professionally and academically, and his work often incited my enthusiasm for public finance. In addition, I received many helpful comments from Pierre Boyer, Emanuel Hansen, Aleh Tsyvinski and Nicolas Werquin.

I benefited from financial and administrative support from the Bonn Graduate School of Economics, the Max Planck Institute for Research on Collective Goods, Yale University, and Harvard Business School. I would like to thank Silke Kinzig, Monika Stimpson, Heidi Morgenstern and Pamela O'Donnell for all their help. Further, I thank my classmates Thomas Graeber, Fabian Walders, Antonis Kotidis and Daxin Dong who have made my studies much more enjoyable.

Finally, I am forever grateful to my parents Friedrich Georg and Marianne Scherf and my sister Miriam Scherf who have supported me unconditionally throughout my studies. Most importantly, I am deeply thankful to my wife Alexandra Scherf for bringing countless blessings to my life and always believing in me.

Contents

Preface

This dissertation consists of three self-contained essays on public finance, that is, on the study of government revenue and expenditure. Taxation and public spending reflect society's values, priorities, as well as its political and informational constraints. This makes them a fascinating subject for economic research and a frequent topic of political debate.

Economists have recently studied the optimal design of tax policy using the Mirrleesian mechanism design framework. At the core of this analysis lies a trade-off between equity and efficiency: In order to maximize social welfare, a utilitarian social planner wishes to redistribute consumption from individuals who have high income-earning ability to those with less income-earning ability. Because individual abilities are unobserved by the social planner, the optimal tax system must be incentive compatible and involves distortions of labor efforts. Chapters 1 and 2 of this thesis contain analyses of this trade-off and the resulting labor distortions. Researchers have also considered the effects of political competition on income taxes and public good provision. This part of the literature assumes that politicians design tax schedules in order to win elections instead of maximizing social welfare. Chapters 1 and 3 concern the political economy of taxation and public good provision. Methodologically, the analyses in all three chapters utilize tools from microeconomic theory, in particular mechanism design and game theory.

Chapter 1 jointly studies public good provision and nonlinear income taxation. It argues that public revenue and spending should be analyzed simultaneously, because the two are interdependent. Specifically, it assumes that individual income earning ability is a function of innate talent and a public good. The public good in turn is financed by distortionary taxes. The chapter then studies how public good provision affects the efficiency of optimal income taxes and, conversely, how taxation affects the efficiency of public good provision. It first characterizes Pareto efficient allocations and an allocation that results from political competition over both public good provision and nonlinear income taxation. In the political equilibrium, the median voter's favorite policy is the Condorcet winner in an election. Further, it shows that public good provision helps to mitigate the incentive problems of income taxation and that the efficiency of public good provision is dependent on the position of the median voter in the income distribution.

Chapter 2 studies optimal income taxation when individuals exhibit intention-based preferences. The chapter is based on the conventional income taxation model with individuals who differ in their skill type and a social planner who redistributes income subject to incentive compatibility and resource constraints. However, it augments the standard taxation mechanism by offering low-ability types the choice of exerting a higher labor effort, which results in a higher utility for high-ability types. Under intention-based preferences, this creates slack in the incentive constraint for high-ability types and allows us to implement an allocation that is Pareto superior to the allocation that results from the standard mechanism. The interpretation is that the rich do not mind helping the poor if they "do their part" by working harder. Interestingly, with intention-based preferences the optimal allocation lies outside the Pareto frontier of the standard model and relies on individuals' concern for procedural justice, which cannot be characterized as the maximization of a social welfare function.

Chapter 3 characterizes demand for think tanks in the presence of academic experts. It poses the question: how can demand for think tanks co-exist with research that is made publicly available by academic researchers? To investigate this question the chapter proposes a model in which voters differ in their pre-tax incomes and exhibit uncertainty about the deadweight cost caused by taxation. For example, they might be uncertain about the elasticity of taxable income or the administrative cost of government. An academic expert observes the true deadweight cost of the tax system and communicates this information to the electorate via a cheap talk message. Additionally, voters can choose to pool their resources to finance a partisan think tank that has access to the same information as the academic expert. The chapter shows that individuals are willing to pay for a think tank if and only if the academic expert is partisan with probability greater than zero. That is, demand for think tanks exists if and only if academic experts represent the interests of a certain part of the electorate, instead of always providing truthful information.

Chapter 1

Voting over Public Good Provision and Nonlinear Income Tax Schedules

1.1 Introduction

In the public finance literature, theories of optimal income taxation and theories of public good provision have mostly been treated separately. This is problematic, because income taxation contributes to the financing of public goods which implies clear interdependencies between the two. Further, public goods may impact the incentive problems of income taxation and thus influence individual preferences over taxes. Optimal policies of income taxation and public good provision are thus affected by their interdependency and should be analyzed jointly. This becomes especially obvious when public goods affect individuals' income earning abilities.

This paper develops a model of nonlinear taxation in which taxes not only serve to redistribute income but also to finance a public good. Individual income earning ability is a function of the public good and innate ability, which differs across individuals. The paper characterizes Pareto optimal public good provision and income taxataion, subject to incentive compatibility and resource constraints. It thereby addresses two main questions: How does public good provision affect the incentive problems of income taxation? And how does this retroact on Pareto efficient public good provision? After analyzing Pareto efficient policies, the paper then considers the income tax schedule and public good provision that arise out of political competition.

Related literature

The workhorse model of nonlinear income taxation is Mirrlees (1971). I extend the standard model in two ways. First, while the standard model assumes that tax revenue is used only for redistribution, this paper allows for investments of tax revenue into a public good that increases the income earning abilities of individuals. Second, in addition to analyzing Pareto optimal taxation, this paper characterizes tax schedules that arise out of political competition.

Political competition over linear tax schedules has been studied by Roberts (1977) and Meltzer and Richard (1981). Recently, nonlinear income taxation has been connected with political economy by Bierbrauer and Boyer (2013) and Bierbrauer and Boyer (2015), who study Downsian competition over nonlinear tax schedules. This paper is most closely related to Brett and Weymark (2017) and Roell (2012), who analyze nonlinear taxation in a citizen-candidate framework of political competition. I adopt their characterization of political competition over tax schedules, but expand the policy space by allowing investments of tax revenue into public good provision.

The preference model of individual income earning ability as a function of innate talent and public goods is adopted from Weinzierl (2014), who characterizes optimal benefit-based taxation when individuals' benefits from public goods are given by their income earning abilities. Matsumoto (2001) also considers productivity enhancing public goods and analyzes their effect on the incentive problems of taxation. Conversely, Atkinson and Stern (1974) and Boadway and Keen (1993) have studied the optimal provision of public goods that are financed with distortionary taxes.

1.2 The model

There is an economy with n citizens who differ in their innate ability types $a_1, ..., a_n$, with $a_j > a_k \Leftrightarrow j > k$. Utility for a citizen of type a_i is given by

$$u_i(c_i, y_i, G) \;=\; c_i - v\left(\frac{y_i}{f(a_i, G)}\right) \tag{1.1}$$

where y_i and c_i denote his pre- and after tax income, respectively. There is a disutility of labor $v(\cdot)$ with $v'(\cdot) > 0$ and $v''(\cdot) > 0$. As in Weinzierl (2014), citizen i's productivity is given by a function $f(a_i, G)$ which depends on his innate ability a_i and the provided level of a public good G. Individual productivity increases in both a_i and G, but at a diminishing rate, that is $f_G > 0$, $f_a > 0$, $f_{G,G} < 0$, and $f_{a,a} < 0$, which is an intuitive assumption for a production function.

In this model, individuals derive no direct utility from the public good. Instead, public good provision increases the income earning abilities of individuals. Possible examples of such goods could be public transportation, or investments in communication infrastructure such as broadband internet. These goods do not bring direct pleasure to the people who use them, but increase their welfare by enabling them to work more efficiently.

Note that the specifications of $f(a_i, G)$ imply that the Spence-Mirrlees single-crossing property holds, independent of public good provision, since

$$\frac{\partial}{\partial a_i} \frac{v'\left(\frac{y_i}{f(a_i,G)}\right)}{f(a_i, G)} = -\frac{f_a(a_i, G)}{f(a_i, G)} \left(\frac{v''\left(\frac{y_i}{f(a_i,G)}\right)}{f(a_i, G)} y_i + v'\left(\frac{y_i}{f(a_i, G)}\right)\right) < 0 \quad (1.2)$$

Hence, in an incentive compatible tax schedule, individuals of a higher ability type will always be "richer" than individuals of a lower ability type in the sense of pre- and after-tax income. As in Weinzierl (2014), I make use of the following definitions and additional assumption on the shape of $f(a_i, G)$:

Assumption 1: *The production function is multiplicative, that is* $f(a_i, G) = h(a_i)g(G)$ *for two differentiable functions* $h(a_i)$ *and* $g(G)$*, both* $R^+ \rightarrow R^+$*.*

Note that the commonly used Cobb-Douglas production function is a special case of this multiplicative functional form.

Definition 1: *The elasticity of individual productivity with respect to public good provision,* $\varepsilon^G(G)$*, is defined as:*

$$\varepsilon^G(G) = \frac{f_G(a_i, G)}{f(a_i, G)} G = \frac{g'(G)}{g(G)} G \quad (1.3)$$

$\varepsilon^G(G)$ denotes the percentage increase of $f(a_i, G)$, if G increases by one percent. Assumption 1 is in fact equivalent to this elasticity being constant across all individuals. The standard model can be seen as a special case of this assumption, namely where $\varepsilon^G(G) = 0$.

1.2.1 Individually optimal policies

A policy consists of a tax system and a level of public good provision. The public good is financed with tax revenue and has unit cost p. Admissible policies have to be incentive compatible and budget balanced. By the taxation principle, a full description of a tax system will be given by a list of pre- and after tax incomes for all types. Hence, a policy $\{(c_j, y_j)_{j=1}^n, G\}$ consists of a list of pre-tax and after-tax incomes for all types and a level of public good provision.

As a first step, let me describe the first-best policy that maximizes Utilitarian social welfare under full information over individuals' types.

Lemma 1: *A first-best policy* $\{(c_j^{FB}, y_j^{FB})_{j=1}^n, G^{FB}\}$ *is characterized by the following three first-order conditions:*

$$v'\left(\frac{y_j^{FB}}{f(a_j, G^{FB})}\right)\frac{1}{f(a_j, G^{FB})} = 1 \quad \forall j = 1, ..., n \tag{1.4}$$

$$\varepsilon^G(G^{FB})\sum_{j=1}^n y_j^{FB} = pG^{FB} \tag{1.5}$$

$$\sum_{j=1}^n \left(y_j^{FB} - c_j^{FB}\right) - pG^{FB} = 0 \tag{1.6}$$

First-best policies are characterized by three equations. First, labor efforts y_j are elicited efficiently if the marginal cost of labor equals the marginal benefit of consumption. Second, efficient public good provision is described by the Samuelson rule, namely that the sum of citizens' marginal benefits of public good provision should equal the marginal cost of the public good. Third, the public budget should be balanced.

Next, let me characterize Pareto optimal policies in second-best, that is, when ability types are private information. Specifically, as the extreme cases of Pareto efficient allocations, I characterize individually optimal policies. A policy is individually optimal if it maximizes the utility of one citizen i, subject to incentive compatibility and budget constraints. For the remainder of this section, let $i \in \{1, ..., n\}$ denote the citizen for whom the second-best policy is individually optimal. Formally, the policy solves the following maximization problem:

$$\max_{(c_1,...,c_n,y_1,...,y_n,G)} u_i(c_i, y_i, G) \tag{1.7}$$

subject to

$$u_j(c_j, y_j, G) \geq u_j(c_{j+1}, y_{j+1}, G) \quad \forall j = 1, ..., n-1 \tag{1.8}$$

$$u_j(c_j, y_j, G) \geq u_j(c_{j-1}, y_{j-1}, G) \quad \forall j = 2, ..., n \tag{1.9}$$

$$\sum_{j=1}^n (y_j - c_j) - pG \geq 0 \tag{1.10}$$

Note that single-crossing implies that incentive compatibility constraints are redundant for non-adjacent types (Roell, 2012). The following two lemmas about the individually optimal tax schedule of citizen i are adopted from Roell (2012).

Lemma 2: *At the optimum, the budget constraint is always binding:*

$$\sum_{j=1}^{n}(y_j - c_j) - pG = 0 \tag{1.11}$$

To see this, note that if the budget constraint was slack by some $\varepsilon > 0$, then there exists a policy $\{(c_j + \frac{\varepsilon}{n}, y_j)_{j=1}^{n}, G\}$ that is incentive compatible, satisfies the budget constraint, and gives citizen i higher utility. Hence the original tax schedule could not have been individually optimal for i.

Lemma 3: *For all* $j < i$*, the bundles* (c_j, y_j) *and* (c_{j+1}, y_{j+1}) *are connected by citizen* j*'s upward incentive constraint. For all* $j > i$*, the bundles* (c_j, y_j) *and* (c_{j-1}, y_{j-1}) *are connected by citizen* j*'s downward incentive constraint. That is:*

$$u_j(c_j, y_j, G) = u_j(c_{j+1}, y_{j+1}, G) \quad \forall j < i \tag{1.12}$$

$$u_j(c_j, y_j, G) = u_j(c_{j-1}, y_{j-1}, G) \quad \forall j > i \tag{1.13}$$

Making use of these lemmas and maximizing (1.7) subject to (1.11), (1.12), (1.13), the following are the first-order conditions with respect to $(y_1, ..., y_n)$:

First-order condition with respect to y_j, $j < i$:

$$\begin{aligned} \phi_j^M := \; & 1 - v'\left(\frac{y_j}{f(a_j, G)}\right)\frac{1}{f(a_j, G)} \\ + \; & (j-1)\left[v'\left(\frac{y_j}{f(a_{j-1}, G)}\right)\frac{1}{f(a_{j-1}, G)} - v'\left(\frac{y_j}{f(a_j, G)}\right)\frac{1}{f(a_j, G)}\right] \\ = \; & 0 \end{aligned} \tag{1.14}$$

First-order condition with respect to y_j, $j > i$:

$$\begin{aligned} \phi_j^R := \; & 1 - v'\left(\frac{y_j}{f(a_j, G)}\right)\frac{1}{f(a_j, G)} \\ + \; & (n-j)\left[v'\left(\frac{y_j}{f(a_{j+1}, G)}\right)\frac{1}{f(a_{j+1}, G)} - v'\left(\frac{y_j}{f(a_j, G)}\right)\frac{1}{f(a_j, G)}\right] \\ = \; & 0 \end{aligned} \tag{1.15}$$

First-order condition with respect to y_i:

$$\begin{aligned}\phi_i \;:=\; & 1 - v'\left(\frac{y_i}{f(a_i, G)}\right)\frac{1}{f(a_i, G)} \\ + \;& (i-1)\left[v'\left(\frac{y_i}{f(a_{i-1}, G)}\right)\frac{1}{f(a_{i-1}, G)} - v'\left(\frac{y_i}{f(a_i, G)}\right)\frac{1}{f(a_i, G)}\right] \\ + \;& (n-i)\left[v'\left(\frac{y_i}{f(a_{i+1}, G)}\right)\frac{1}{f(a_{i+1}, G)} - v'\left(\frac{y_i}{f(a_i, G)}\right)\frac{1}{f(a_i, G)}\right] \\ = \;& 0\end{aligned} \tag{1.16}$$

Brett and Weymark (2017) call ϕ_j^M and ϕ_j^R maxi-max and Rawlsian tax schedules, respectively. The reason for this choice of names is that in order to maximize citizen i's utility, resources are diverted upwards from types below her (as in a "maxi-max" tax schedule that maximizes the utility of the highest type) and downwards from types above her (as in a Rawlsian tax schedule that maximizes the utility of the lowest type).

To interpret the shape of the tax schedules, let us first consider ϕ_j^M and recall that for all $j < i$, consumption-labor bundles are connected by the upward incentive constraints. ϕ_j^M describes the marginal benefit for citizen i when the labor effort of some citizen j is increased by one unit while maintaining incentive compatibility and budget constraints. Increasing citizen j's labor effort creates one unit of consumption that can be diverted to citizen i. By compensating j with $v'\left(\frac{y_j}{f(a_j,G)}\right)\frac{1}{f(a_j,G)}$ units of consumption, he stays on his initial indifference curve and has no incentive to mimick the type above him. This move of j up his indifference curve also leaves the incentive constraints of types above j untouched. However, the (upward) incentive constraint of citizen $j-1$ is now slack. Hence i can extract from all types below j the amount of consumption that will make citizen $j-1$ indifferent between his old bundle and j's new bundle. The sum of these three effects makes up ϕ_j^M. (Brett and Weymark, 2017).

Similarly, ϕ_j^R describes the marginal benefit for citizen i of increasing the labor effort of some citizen $j > i$. The difference to ϕ_j^M is that for types above i it is downwards incentive constraints that bind. Moving a citizen $j > i$ up his indifference curve as described above violates the (downward) incentive constraints for citizen $j+1$. Hence, all $n-j$ types above j need to be compensated in terms of consumption in order to restore incentive compatibility.

There is an important difference between these tax schedules and the ones characterized in Brett and Weymark (2017). In their model, ϕ_j^M and ϕ_j^M are independent of the type of citizen i, that is, the same amount of labor is elicited from citizen j, no matter if the policy is individually optimal for citizen i or for another citizen k (with $i, k \neq j$). Here, on the other hand, ϕ_j^M and ϕ_j^R are technically functions of G. Thus, the amount of labor that is elicited from citizen j depends on the type of citizen i, if different types prefer different levels of public good provision.

Given the amount of the public good provided, effort of types below i will be distorted upward, whereas effort is distorted downward for types above i. Only types 1's and n's effort levels are undistorted. The implicit marginal tax rate facing citizen j, defined as

$$\tau_j := 1 - v'\left(\frac{y_j}{f(a_j, G)}\right)\frac{1}{f(a_j, G)} \tag{1.17}$$

is such that

$$\begin{cases} \tau_j = 0 & , \text{if } j \in \{1, n\} \\ \tau_j < 0 & , \text{if } 1 < j < i \\ 0 < \tau_j < 1 & , \text{if } i < j < n \end{cases} \tag{1.18}$$

Proposition 1: *Public good provision reduces the labor distortions for every citizen. That is, as* G *increases,* τ_j *gets closer to zero for all* $j = 2, \dots, n-1$.

The public good helps to mitigate the incentive problems in the tax schedule and reduces labor distortions. Labor distortions arise because of incentive compatibility constraints. Increasing the amount of public good provision reduces the extent to which additional labor effort creates slack in the upward incentive constraint. Since labor efforts for types below i are distorted upwards, the public good reduces the distortion by making it less attractive for i to elicit additional labor effort from types below him. Similarly, the public good reduces the extent to which additional labor effort violates the downward binding incentive constraints. Types above i, who's labor efforts are distorted downwards, thus require less compensation for incentive compatibility to be restored. The public good thus allows i to elicit a higher labor effort from higher types and reduces their labor distortions as well.

The first-order condition of citizen i's maximization problem with respect to G is given by:

$$\begin{aligned} p \; = \; & \sum_{j=1}^{n} v'\left(\frac{y_j}{f(a_j,G)}\right)\frac{f_G(a_j,G)}{f(a_j,G)^2}y_j & (1.19) \\ + \; & \sum_{j=1}^{i}(j-1)\left[v'\left(\frac{y_j}{f(a_j,G)}\right)\frac{f_G(a_j,G)}{f(a_j,G)^2}y_j - v'\left(\frac{y_j}{f(a_{j-1},G)}\right)\frac{f_G(a_{j-1},G)}{f(a_{j-1},G)^2}y_j\right] \\ + \; & \sum_{j=i}^{n}(n-j)\left[v'\left(\frac{y_j}{f(a_j,G)}\right)\frac{f_G(a_j,G)}{f(a_j,G)^2}y_j - v'\left(\frac{y_j}{f(a_{j+1},G)}\right)\frac{f_G(a_{j+1},G)}{f(a_{j+1},G)^2}y_j\right] \\ \Leftrightarrow \; pG \; = \; & \varepsilon^G(G)\sum_{j=1}^{n} y_j & (1.20) \end{aligned}$$

Equation (1.19) depicts the first-order condition in a way similar to the characterization by Boadway and Keen (1993). On the left hand side there is the marginal cost of public good provision p. The first term on the right hand side equals the sum of individuals' marginal utilities of public good provision. The Samuelson rule would end here. The additional two terms capture the effect that the public good has on the incentive constraints of the tax schedule. The first term concerns the upwards binding constraints for all types below i's, and the second term the downwards binding constraints for all types above i's. Note that these two terms are not zero, as the weak separability condition of Boadway and Keen (1993) is not given here, and thus the first-order condition does not equal the Samuelson rule. That is, citizen i's individually optimal policy may include a level of public good provision that is inefficient given the amount of labor elicited.

There are two effects at work here that influence citizen i's optimal level of G. The first effect is that the public good increases i's productivity and hence her utility. It also increases everybody else's utility and thus creates additional room to divert resources away from all other citizens. This direct effect thus unambiguously increases the citizen i's optimal level of G. Note that the utility gain is increasing in the type of citizen i. Financing the increase in G requires eliciting higher amounts of labor from everyone. The higher the type of citizen i, the lower her disutility from labor and thus the perceived cost of an extra unit of G. Further, $\varepsilon^G(G)$ being constant across types implies that the public good gives higher types a greater marginal increase in income earning ability than lower types.

The second effect at work here is the indirect effect that G has on the incentive constraints. For all citizens above i, that is on the downwards-binding part of the tax schedule, any increase in G relaxes the incentive constraints and gives room to divert resources away from them towards i. For all citizens below i, however, incentive constraints are upwards-binding and hence tightened by increases in G. Any extra unit of G that is provided thus requires compensating all citizens below i in order to restore incentive compatibility. This second effect hence increases the optimal level of G if citizen i is of a relatively low type, but decreases the optimal level of G if i of a relatively high type.

Substituting in equations (1.14), (1.15), and (1.16), the optimality condition simplifies to (1.20). Equation (1.20) has a very clear interpretation: no matter for which type a policy is individually optimal, it is always best to invest a share $\varepsilon^{G}(G)$ of the total output produced into public good provision. Hence a fraction $\left(1 - \varepsilon^{G}(G)\right)$ of total output is then used for redistribution in the form of consumption, as can easily be seen when combining equation (1.20) with the public budget constraint:

$$\sum_{j=1}^{n} c_j \;=\; \left(1 - \varepsilon^{G}(G)\right) \sum_{j=1}^{n} y_j \tag{1.21}$$

Even though equation (1.20) does not lead to first-best efficient provision of the public good, as explained above, it has in fact the same mathematical structure as the first-best Samuelson rule (1.5). It turns out that when the public good is financed by labor incomes that are distorted through a nonlinear income schedule, the distortions of labor efforts cancel out some of the distortions of public good provision. This is because public good provision and taxation have opposite effects on the incentive constraints. An increase in G relaxes downwards-binding incentive constraints and tightens upwards-binding constraints, while the reverse is true for eliciting additional labor efforts.

1.2.2 Political equilibrium

I can now proceed to characterize the political equilibrium. In the following subsection, I use superscripts to denote the identity of the individual for whom a policy is individually optimal. That is, let $\{(c_j^i, y_j^i)_{j=1}^n, G^i\}$ denote the individually optimal policy for citizen i.

As in Brett and Weymark (2017), every citizen proposes their individually optimal policy and all citizens engage in pairwise majority voting over the proposed policies. Thus, if a Condorcet winner exists she is elected as policy maker and implements her individually optimal policy. The timing works as follows: First, citizens propose their favorite policies and the Condorcet winner is elected as policy maker. Second, each citizen chooses his pre- and after tax income from the policy maker's proposed schedule, and the policy maker invests the proposed amount of tax revenue into the public good. Third, utilities are realized.

Proposition 2: *Citizens of a higher ability type propose greater investments into the public good:*

$$G^i < G^{i+1} \; \forall i = 1, ..., n-1 \tag{1.22}$$

Recall the two effects that are at work here. Increasing G increases everyone's income earning ability and hence creates more resources that can be diverted towards the policy maker. Citizens of higher types do not mind so much the extra labor effort it takes to finance the public good. Further, because of the positive complementartiy of public goods and innate talents, higher types derive higher marginal utility from an extra unit of G. However, increasing G also means having to compensate lower types with more consumption, because public good provision tightens the upwards incentive constraints. Overall, the first, direct, effect outweighs the cost of restoring incentive compatibility, and hence citizens of higher types propose greater investments into public good provision.

Since we know that $v'\left(\frac{y_j}{f(a_j,G)}\right)\frac{1}{f(a_j,G)} > 1$ on the maxi-max tax-schedule, it becomes clear that citizens of high types, who's individually optimal tax schedules are mostly maxi-max, propose a higher level of G than the first-best Samuelson rule suggests. Citizens of low types, on the other hand, propose lower levels of G than the first-best Samuelson rule suggests.

Corollary 1: *Citizens of a low type propose levels of public good provision that are lower than the first-best level, whereas citizens of a high type propose levels that are higher than the first-best level:*

$$G^1 < G^{FB} < G^n \tag{1.23}$$

We know that for the lowest type it holds that $\sum_{j=1}^{n} y_j^1 < \sum_{j=1}^{n} y_j^{FB}$. It then follows from (1.5) and (1.20) that $G^1 < G^{FB}$. The reverse is true for the highest type. Based on this corollary I conjecture that, if the type space is not too coarse, there exists a citizen who proposes the first-best level of public good provision.

Lemma 4: *Citizens of a higher ability type propose to elicit a higher effort from all citizens:*

$$y_j^i < y_j^{i+1} \; \forall j = 1, ..., n-1 \tag{1.24}$$

This observation is intuitive, since higher types are more productive and thus more willing to exert a higher labor effort. Further, Proposition 2 states that higher types propose higher levels of public good provision. Increasing the provision of G while maintaining incentive compatibility further increases the need for additional labor efforts by the citizens.

Proposition 3: *All citizens have single-peaked preferences over the individually optimal policies* $\{(c_j^i, y_j^i)_{j=1}^n, G^i\}_{i=1}^n$*, that is:*

$$u_j(c_j^i, y_j^i, G^i) \;\leq\; u_j(c_j^{i+1}, y_j^{i+1}, G^{i+1}) \quad \forall j \;>\; i \tag{1.25}$$
$$u_j(c_j^i, y_j^i, G^i) \;\geq\; u_j(c_j^{i+1}, y_j^{i+1}, G^{i+1}) \quad \forall j \;<\; i \tag{1.26}$$

Thus, there exists a Condorcet winner, namely the citizen of type a_M*, where* a_M *is the median of* $\{a_1, ..., a_n\}$.

This result shows that citizens have single-peaked preferences over individually optimal tax schedules even when tax revenue is used to finance a public good. It implies that in the voting process described above, a Condorcet winner always exists and that she is the citizen of the median type. The fact that the policy that arises in political equilibrium is individually optimal for the median type provides support for Director's Law, as it can be interpreted as redistribution towards the middle class (Brett and Weymark, 2017).

1.3 Conclusion

The paper suggests that income taxation and public good provision ought to be analyzed jointly, as they exhibit clear interdependencies when taxes do not only serve to redistribute income, but also finance public spending. Modeling income earning ability as a function of innate talent and public good provision makes this evident.

I first characterized individually optimal policies as the extreme cases of Pareto efficient allocations and showed that public good provision mitigates the incentive problems of income taxation. Then I characterized the political equilibrium for this model by showing that individuals have single-peaked preferences over individually optimal policies and thus the median voter's favorite policy is the Condorcet winner in election. We see that the level of public good provision in political equilibrium is in fact increasing in the type of the median voter. That is, the higher up the median voter stands in the income distribution, the more public spending we should expect. We also see that public good provision is inefficiently low (compared to the first-best level) if the median voter's position in the income distribution is relatively low. Conversely, public good provision is inefficiently high if the median voter is high up in the income distribution. Thus, efficiency of public spending depends on the skewdness of the income distribution.

Appendix 1.A

Proof of Lemma 1:

Maximizing the utilitarian social welfare function

$$\max_{(c_1,\ldots,c_n,y_1,\ldots,y_n,G)} \sum_{j=1}^{n} \left(c_j - v\left(\frac{y_j}{f(a_j, G)}\right)\right)$$

subject to the public budget constraint

$$\sum_{j=1}^{n} (y_j - c_j) - pG \;=\; 0$$

yields the first-order conditions as stated by Lemma 1.

Proofs of Lemma 2 and Lemma 3:

see Roell (2012)

Proof of Proposition 1:

Consider first the maxi-max tax schedule ϕ_j^M in equation (1.14). The distortion of citizen j's labor supply is given by the term

$$v'\left(\frac{y_j}{f(a_{j-1}, G)}\right)\frac{1}{f(a_{j-1}, G)} - v'\left(\frac{y_j}{f(a_j, G)}\right)\frac{1}{f(a_j, G)} \;>\; 0 \tag{1.27}$$

If we increase G, this distortionary term becomes smaller since

$$\frac{\partial}{\partial G}\left[v'\left(\frac{y_j}{f(a_{j-1},G)}\right)\frac{1}{f(a_{j-1},G)} - v'\left(\frac{y_j}{f(a_j,G)}\right)\frac{1}{f(a_j,G)}\right] \tag{1.28}$$

$$= \varepsilon^G(G)\Big[v''\left(\frac{y_j}{f(a_j,G)}\right)\frac{y_j}{f(a_j,G)^2} - v''\left(\frac{y_j}{f(a_{j-1},G)}\right)\frac{y_j}{f(a_{j-1},G)^2}$$
$$+v'\left(\frac{y_j}{f(a_j,G)}\right)\frac{1}{f(a_j,G)} - v'\left(\frac{y_j}{f(a_{j-1},G)}\right)\frac{1}{f(a_{j-1},G)}\Big] \;<\; 0 \tag{1.29}$$

Conversely, the distortionary term in the Rawlsian tax schedule ϕ_j^R, which is negative, is increasing in G.

Proof of Proposition 2:

Making use of Assumption 1, the FOC with respect to public good provision (1.19) is given by:

$$\frac{\varepsilon^G(G^i)}{G^i}\sum_{j=1}^{n} y_j^i - p \;=\; 0 \tag{1.30}$$

Using (1.14), (1.15), and (1.16) we get that

$$\frac{\partial}{\partial a_i}\left[\frac{\varepsilon^G(G^i)}{G^i}\sum_{j=1}^{n} y_j^i - p\right] \;=\; \frac{\varepsilon^G(G^i)}{G^i}\frac{\partial y_i}{\partial a_i} \;>\; 0 \tag{1.31}$$

Further, by optimality of $\{(c_j^i, y_j^i)_{j=1}^n, G^i\}$ it needs to hold that

$$\frac{\partial}{\partial G}\left[\frac{\varepsilon^G(G^i)}{G^i}\sum_{j=1}^{n} y_j^i - p\right] \;<\; 0 \tag{1.32}$$

so that the implicit function theorem implies that

$$\frac{\partial G^i}{\partial a_i} \;=\; -\frac{\frac{\partial}{\partial a_i}\left[\frac{\varepsilon^G(G^i)}{G^i}\sum_{j=1}^{n} y_j^i - p\right]}{\frac{\partial}{\partial G}\left[\frac{\varepsilon^G(G^i)}{G^i}\sum_{j=1}^{n} y_j^i - p\right]} \;>\; 0 \tag{1.33}$$

Proof of Lemma 4:

Consider the case $j < i$. By optimality of $\{(c_j^i, y_j^i)_{j=1}^n, G^i\}$ it needs to hold that

$$\frac{\partial \phi_j^M}{\partial y_j^i} \;<\; 0 \tag{1.34}$$

Making use of Assumption 1 and simplifying, we get that

$$\frac{\partial \phi_j^M}{\partial G} \;=\; -\frac{\varepsilon^G(G^i)}{G^i}\left(y_j^i\frac{\partial \phi_j^M}{\partial y_j^i} - 1\right) \;>\; 0 \tag{1.35}$$

so that applying the implicit function theorem and simplifying yields

$$\frac{\partial y_j^i}{\partial G} \;=\; -\frac{\frac{\partial \phi_j^M}{\partial G}}{\frac{\partial \phi_j^M}{\partial y_j^i}} \;=\; \frac{\varepsilon^G(G^i)}{G^i}\left(y_j^i - \frac{1}{\frac{\partial \phi_j^M}{\partial y_j^i}}\right) \;>\; 0 \tag{1.36}$$

By the same logic this can be shown for $j \geq i$. The result then follows, since $\frac{\partial y_j^i}{\partial a_i} = \frac{\partial y_j^i}{\partial G^i}\frac{\partial G^i}{\partial a_i} > 0$.

Proof of Proposition 3:

Before beginning the main proof I first show that for all $j < i$ it holds that

$$\frac{\partial}{\partial a_i}\left(\frac{y_j^i}{g(G^i)}\right) > 0 \tag{1.37}$$

$$\begin{aligned}
\Leftrightarrow\quad & \frac{\frac{\partial y_j^i}{\partial a_i} g(G^i) - g'(G^i)\frac{\partial G^i}{\partial a_i} y_j^i}{g(G^i)^2} > 0 && (1.38)\\
\Leftrightarrow\quad & \frac{\partial y_j^i}{\partial a_i} g(G^i) - g'(G^i)\frac{\partial G^i}{\partial a_i} y_j^i > 0 && (1.39)\\
\Leftrightarrow\quad & \frac{\partial y_j^i}{\partial a_i} - \frac{g'(G^i)}{g(G^i)}\frac{\partial G^i}{\partial a_i} y_j^i > 0 && (1.40)\\
\Leftrightarrow\quad & \frac{\partial y_j^i}{\partial G^i}\frac{\partial G^i}{\partial a_i} - \frac{g'(G^i)}{g(G^i)}\frac{\partial G^i}{\partial a_i} y_j^i > 0 && (1.41)\\
\Leftrightarrow\quad & \frac{\partial y_j^i}{\partial G^i} - \frac{\varepsilon^G(G^i)}{G^i} y_j^i > 0 && (1.42)\\
\Leftrightarrow\quad & \frac{\varepsilon^G(G^i)}{G^i}\left(y_j^i - \frac{1}{\frac{\partial \phi_j^M}{\partial y_j^i}}\right) - \frac{\varepsilon^G(G^i)}{G^i} y_j^i > 0 && (1.43)\\
\Leftrightarrow\quad & -\frac{\varepsilon^G(G^i)}{G^i}\frac{1}{\frac{\partial \phi_j^M}{\partial y_j^i}} > 0 && (1.44)\\
& && (1.45)
\end{aligned}$$

The last statement is clearly true since $\frac{\partial \phi_j^M}{\partial y_j^i} < 0$. It then follows that

$$\begin{aligned}
\Rightarrow\quad & \frac{y_j^{i-1}}{f(a_j, G^{i-1})} \le \frac{y_j^i}{f(a_j, G^i)} && (1.46)\\
\Leftrightarrow\quad & \frac{y_j^{i-1}}{f(a_j, G^{i-1})}\frac{f(a_{j-1}, G^{i-1}) - f(a_j, G^{i-1})}{f(a_{j-1}, G^{i-1})} \le \frac{y_j^i}{f(a_j, G^i)}\frac{f(a_{j-1}, G^{i-1}) - f(a_j, G^{i-1})}{f(a_{j-1}, G^{i-1})} && (1.47)\\
\Leftrightarrow\quad & \frac{y_j^{i-1}}{f(a_j, G^{i-1})} - \frac{y_j^{i-1}}{f(a_{j-1}, G^{i-1})} \le \frac{y_j^i}{f(a_j, G^i)} - \frac{y_j^i\, f(a_j, G^{i-1})}{f(a_j, G^i) f(a_{j-1}, G^{i-1})} && (1.48)\\
\Leftrightarrow\quad & \frac{y_j^{i-1}}{f(a_j, G^{i-1})} - \frac{y_j^{i-1}}{f(a_{j-1}, G^{i-1})} \le \frac{y_j^i}{f(a_j, G^i)} - \frac{y_j^i}{f(a_{j-1}, G^i)} && (1.49)
\end{aligned}$$

Where the last step follows because Assumption 1 implies that

$$\frac{f(a_j, G^{i-1})}{f(a_j, G^i)} = \frac{f(a_{j-1}, G^{i-1})}{f(a_{j-1}, G^i)} \tag{1.50}$$

Finally, by convexity of $v(\cdot)$ and Lemma 4 it follows that

$$v\left(\frac{y_j^{i-1}}{f(a_j, G^{i-1})}\right) - v\left(\frac{y_j^{i-1}}{f(a_{j-1}, G^{i-1})}\right) \ge v\left(\frac{y_j^i}{f(a_j, G^i)}\right) - v\left(\frac{y_j^i}{f(a_{j-1}, G^i)}\right) \forall j < i \tag{1.51}$$

Similarly, for all $j > i$ it holds that

$$v\left(\frac{y_j^{i+1}}{f(a_j, G^{i+1})}\right) - v\left(\frac{y_j^{i+1}}{f(a_{j+1}, G^{i+1})}\right) \geq v\left(\frac{y_j^i}{f(a_j, G^i)}\right) - v\left(\frac{y_j^i}{f(a_{j+1}, G^i)}\right) \forall j > i \tag{1.52}$$

The proof is then by induction. Consider first the case $j > i$.
Base case (holds by optimality of $\{(c_j^{i+1}, y_j^{i+1})_{j=1}^n, G^{i+1}\}$):

$$u_{i+1}(c_{i+1}^i, y_{i+1}^i, G^i) \leq u_{i+1}(c_{i+1}^{i+1}, y_{i+1}^{i+1}, G^{i+1}) \tag{1.53}$$

Inductive step:

$$u_j(c_j^i, y_j^i, G^i) \leq u_j(c_j^{i+1}, y_j^{i+1}, G^{i+1}) \tag{1.54}$$

together with Lemma 4 and

$$v\left(\frac{y_j^{i+1}}{f(a_j, G^{i+1})}\right) - v\left(\frac{y_j^{i+1}}{f(a_{j+1}, G^{i+1})}\right) \geq v\left(\frac{y_j^i}{f(a_j, G^i)}\right) - v\left(\frac{y_j^i}{f(a_{j+1}, G^i)}\right) \tag{1.55}$$

implies that

$$u_{j+1}(c_j^i, y_j^i, G^i) \leq u_{j+1}(c_j^{i+1}, y_j^{i+1}, G^{i+1}) \tag{1.56}$$

so that by Lemma 3 it follows that:

$$u_{j+1}(c_{j+1}^i, y_{j+1}^i, G^i) = u_{j+1}(c_j^i, y_j^i, G^i) \tag{1.57}$$
$$\leq u_{j+1}(c_j^{i+1}, y_j^{i+1}, G^{i+1}) \tag{1.58}$$
$$= u_{j+1}(c_{j+1}^{i+1}, y_{j+1}^{i+1}, G^{i+1}) \tag{1.59}$$

Conversely, for the case $j < i$:
Base case (holds by optimality of $\{(c_j^{i-1}, y_j^{i-1})_{j=1}^n, G^{i-1}\}$):

$$u_{i-1}(c_{i-1}^i, y_{i-1}^i, G^i) \leq u_{i-1}(c_{i-1}^{i-1}, y_{i-1}^{i-1}, G^{i-1}) \tag{1.60}$$

Inductive step:

$$u_j(c_j^i, y_j^i, G^i) \leq u_j(c_j^{i-1}, y_j^{i-1}, G^{i-1}) \tag{1.61}$$

together with Lemma 4 and

$$v\left(\frac{y_j^{i-1}}{f(a_j, G^{i-1})}\right) - v\left(\frac{y_j^{i-1}}{f(a_{j-1}, G^{i-1})}\right) \geq v\left(\frac{y_j^i}{f(a_j, G^i)}\right) - v\left(\frac{y_j^i}{f(a_{j-1}, G^i)}\right) \tag{1.62}$$

implies that

$$u_{j-1}(c_j^i, y_j^i, G^i) \leq u_{j-1}(c_j^{i-1}, y_j^{i-1}, G^{i-1}) \tag{1.63}$$

so that by Lemma 3 it follows that:

$$u_{j-1}(c_{j-1}^i, y_{j-1}^i, G^i) = u_{j-1}(c_j^i, y_j^i, G^i) \tag{1.64}$$
$$\leq u_{j-1}(c_j^{i-1}, y_j^{i-1}, G^{i-1}) \tag{1.65}$$
$$= u_{j-1}(c_{j-1}^{i-1}, y_{i-1}^{i-1}, G^{i-1}) \tag{1.66}$$

which concludes the proof.

References

Bierbrauer, Felix J., and Pierre C. Boyer. 2013. "Political competition and Mirrleesian income taxation: A first pass." *Journal of Public Economics* 103: 1 –14. [4]

Bierbrauer, Felix J., and Pierre C. Boyer. 2015. "Efficiency, Welfare, and Political Competition." *The Quarterly Journal of Economics* 131 (1): 461–518. [4]

Boadway, Robin, and Michael Keen. 1993. "Public Goods, Self-Selection and Optimal Income Taxation." *International Economic Review* 34 (3): 463–78. [10]

Brett, Craig, and John A. Weymark. 2017. "Voting over selfishly optimal nonlinear income tax schedules." *Games and Economic Behavior* 101 (C): 172–88. [4, 8, 11, 13]

Matsumoto, Mutsumi. 2001. "Public Input Provision in an Optimal Income Tax Model." *FinanzArchiv: Public Finance Analysis* 58 (1): [4]

Meltzer, Allan H., and Scott F. Richard. 1981. "A Rational Theory of the Size of Government." *Journal of Political Economy* 89 (5): 914–27. [4]

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (114): 175–208. [3]

Roberts, Kevin W. 1977. "Voting over income tax schedules." *Journal of Public Economics* 8 (3): 329 –340. [4]

Roell, Ailsa. 2012. "Voting over Nonlinear Income Tax Schedules." Unpublished manuscript. School of International, and Public Affairs, Columbia University. [4, 6, 14]

Weinzierl, Matthew. 2014. "The promise of positive optimal taxation: normative diversity and a role for equal sacrifice." *Journal of Public Economics* 118: 128–42. [4, 5]

Chapter 2

Optimal Income Taxation with Intention-Based Preferences: an Exploration

2.1 Introduction

The human tendency to develop emotional responses to taxation and redistribution is well-documented in the literature on public finance. For example, Bassetto and Phelan (2008) describe "riots" against income taxes and Slemrod (2000) analyzes – in an analogy to Ayn Rand's dystopic novel "Atlas shrugs" – whether high-income earners reduce their labor efforts to protest against taxes that depreciate their contribution to society. Saez and Stantcheva (2016) suggest that so-called "free-loaders" – meaning individuals who choose not to work under an existing transfer system but would choose to work in its absence – are perceived as taking advantage of society. As a consequence, the authors venture, society may wish to disregard the wellbeing of free-loaders in its social welfare objective. Empirical support for this argument can be found in the political success of the earned income tax credit (EITC) in the United States. Arguably, EITC's success stems from its strong work requirement while other programs with weaker work requirements have proven to be politically unpopular. This suggests that people do not mind helping the poor if they work, but the rich may oppose taxes if others do not "do their part."

Although the papers cited above use language that is suggestive of emotional reactions to taxes and redistribution, no work has been done to model such emotions explicitly and to study their effect on optimal tax schedules.

A particularly interesting model to describe such emotions over taxes, in which people wish to reward the poor who work hard and to punish those who "free-load," is given by the intention-based preference model of Rabin (1993). In this model, individuals wish to reciprocate "kindness" of others. That is, individuals like to "help those who are helping them" and "hurt those who are hurting them." (Rabin, 1993)

This paper studies optimal taxation when individuals exhibit intention-based preferences. It utilizes the income tax environment of Stiglitz (1982) where individuals have private information over their skill type and a social planner redistributes income under incentive compatibility and resource constraints. In the welfare maximizing allocation, the incentive constraint for the "high" type is binding and labor effort for the "low" type is distorted downwards. I augment the Stiglitz (1982) mechanism by giving low types the option of exerting a higher labor effort which will increase utility for high types. If individuals exhibit intention-based preferences, high types interpret this additional labor effort as kindness which they wish to reciprocate. This way, intention-based preferences relax the incentive constraint of high types, which allows the social planner to implement an allocation that is Pareto superior to the Stiglitz (1982) allocation. If individuals do not exhibit intention-based preferences, then this mechanism implements the same allocation as Stiglitz (1982).

This result has an important implication for recent attempts to incorporate nonwelfarist objectives into optimal tax theory. Saez and Stantcheva (2016) propose that "generalized social marginal welfare weights" can characterize Pareto optimal tax schedules that take into account nonwelfarist fairness concerns. This paper provides an example of individual fairness concerns, namely intention-based preferences, that make it impossible to characterize Pareto optimal allocations through generalized social welfare weights. Indeed, if individuals exhibit intention-based preferences we can implement allocations that lie outside the Pareto frontier of the standard Stiglitz (1982) model. Further, because these preferences capture a notion of procedural rather than outcome-based justice, we must alter individuals' choice sets in order to implement the optimal allocation.

2.1.1 Related literature

The model draws from the income tax environment of Stiglitz (1982). For a preference model to describe emotions over taxes, I refer to Rabin (1993), which was introduced into the mechanism design framework by Bierbrauer and Netzer (2016). Rabin's model considers individuals who are willing to sacrifice parts of their own payoff in order to reward others who have kind intentions towards them (or to punish those who do not). I augment the Stiglitz (1982) mechanism by allowing less productive individuals to exert a higher labor effort, so that their additional output can be used to increase the consumption of more productive individuals. Highly productive individuals will perceive this additional labor effort as kind and will be willing to support allocations that would violate incentive compatibility in the standard model of Stiglitz (1982). In other words, intention-based preferences create slack in the incentive constraint of "high types." This makes Rabin's model particularly interesting to study the above-mentioned aversion against "free loaders" that was suggested by Saez and Stantcheva (2016).

This paper adapts Rabin's model as a group rule. The literature on group rules, in particular rule utilitarianism, goes back to Harsanyi (1977). It observes that individual behavior often has a negligible effect on overall welfare, thus making it difficult to evaluate individual choices by a Utilitarian metric. Instead, Harsanyi suggests to consider the effect on social welfare that an individual's behavior would have if it became a general rule that was followed by everyone else as well. Harsanyi (1980), Coate and Conlin (2004), and Feddersen and Sandroni (2006) apply this concept to voting behavior, where individuals want to "do their part" in society by following a rule that, if followed by everyone else, would maximize the group's aggregate utility. I use the same logic to formulate intention-based preferences as a group rule. However, instead of being rule Utilitarianists individuals act according to Rabin's model. That is, they evaluate the kindness of an action by considering the consequences that this action would have if it was followed by everybody else.

This paper is related to work on normative diversity and the use of nonwelfarist principles. This literature proposes welfare objectives that are not purely Utilitarian but instead weigh the Utilitarian norm with nonwelfarist objectives. Weinzierl (2014) provides evidence for the prevalence of such normative diversity and its explanatory power for the deviation of existing tax policies from the recommendations of the standard model. Saez and Stantcheva (2016) characterize Pareto efficient allocations by using generalized social welfare weights, which also allow for welfare criteria other than Utilitarianism. Weinzierl (2017) further uses the example of envy to show that rule Utilitarianism can guide public policy when nonwelfarist rules contain relevant information for a welfarist planner with limited information. The current paper differs from this literature by maintaining the Utilitarian criterion to evaluate allocations, but introducing a preference model that captures a procedural (rather than consequentialist) justice concern. This leads to equilibrium allocations that cannot be characterized by a social welfare function or other outcome-based measures such as generalized marginal social welfare weights.

Table 2.1. Prisoners' dilemma

		Player 2	
		C	D
Player 1	C	$\frac{2}{3}, \frac{2}{3}$	$0, 1$
	D	$1, 0$	$\frac{1}{3}, \frac{1}{3}$

The remainder of this paper is organized as follows. I first introduce Rabin's model of intention-based preferences by considering a simple example of a prisoners' dilemma. Second, I outline the income taxation mechanism of Stiglitz (1982). Third, I formulate Rabin (1993) as a group rule and provide a model of intention-based preferences in a large economy, where individual actions have a negligible effect on social welfare. Lastly, I show that with such preferences there exists an allocation that is Pareto superior to the constrained optimum in Stiglitz (1982) and propose a mechanism that implements this allocation as the unique equilibrium.

2.1.2 A simple example of intention-based preferences

This subsection provides a simple example of a prisoners' dilemma to show that under intention-based preferences we can sustain cooperative behavior. This should serve as an explanation of the nature of intention-based preferences, and also give the reader an early illustration of how such preferences can relax individuals' incentive constraints in a model of income taxation.

Consider the following example of a prisoners' dilemma from Rabin (1993). Two players $i = 1, 2$ each choose an action $a_i \in \{C, D\}$: cooperate or defect. Let $\pi_i(a_i, a_j)$ be the "material" payoff for player i if she plays action a_i and her opponent j plays action a_j. Table 2.1 shows the well known matrix of material payoffs for this game. If players only care about their material payoffs, then famously there exists a unique Nash equilibrium in which both players defect.

However, Rabin suggests that if individuals do not only care about material payoffs but also about the "kindness" of each player, then there exists an alternative equilibrium in which both players cooperate. Here, "kindness" refers to a player's effort to increase her opponent's material payoff, irrespective of whether she hurts her own material payoff in the process.

To understand Rabin's argument, suppose it is common knowledge that both players cooperate. Then each player is aware that their opponent sacrifices material payoff equal to $\frac{1}{3}$ in order to help sustain the cooperative outcome. Rabin argues that in choosing to cooperate while knowing that her opponent also cooperates, player i effectively chooses a payoff $\pi_j(C, C) = \frac{2}{3}$ for her opponent, when she could have also chosen to defect and leave her opponent with a lower material payoff of $\pi_j(C, D) = 0$. For this reason, for player i to cooperate means to act kindly towards her opponent. If individuals derive enough utility from reciprocating such kindness, then cooperation may indeed be a mutual best response and (C, C) can be sustained in equilibrium.

To formalize this idea, Rabin proposes that player i's kindness should be measured by how well she leaves her opponent off relative to an "equitable" payoff

$$\pi_j^e(a_i^b) \;=\; \frac{1}{2}\left(\pi_j^h(a_i^b) + \pi_j^l(a_i^b)\right) \tag{2.1}$$

Here, $\pi_j^h(a_i^b)$ denotes the highest material payoff that player j can achieve if he plays some action a_i^b. On the other hand, $\pi_j^l(a_i^b)$ denotes the lowest material payoff that is possible for player j in any of the Pareto efficient outcomes that he can achieve by playing a_i^b. In the prisoners' dilemma, if player j chooses to cooperate, then $\pi_j^h(C) = \frac{2}{3}$ and $\pi_j^l(C) = 0$. His "equitable" payoff is then given by the average of the possible payoffs, namely $\pi_j^e(C) = \frac{1}{3}$.

Rabin's model suggests that player i is "kind" towards her opponent if she leaves player j with a material payoff that is higher than her "equitable" payoff $\pi_j^e(a_i^b)$. Specifically, Rabin defines player i's kindness towards player j as follows.

Definition 1: Player i's kindness towards player j is given by:

$$\kappa_i(a_i, a_j^b) \;=\; \frac{\pi_j(a_i, a_i^b) - \pi_j^e(a_i^b)}{\pi_j^h(a_i^b) - \pi_j^{min}(a_i^b)} \tag{2.2}$$

Here $\pi_j^{min}(a_i^b)$ denotes the worst material payoff that player j can achieve if he plays action a_j^b. It is clear that player i's kindness $\kappa_i(a_i, a_j^b)$ is positive (negative) whenever her action leaves her opponent better (worse) off than the equitable payoff $\pi_j^e(a_i^b)$.

In the prisoners' dilemma example, suppose again player i knows that her opponent will cooperate, that is, $a_i^b = C$. Player j's equitable payoff is then given by $\pi_j^e(a_i^b) = \frac{1}{3}$, the average of the two payoffs that are possible if he cooperates. Thus, if player i chooses to cooperate as well and grant her opponent the higher of the two possible payoffs, then her action results in positive kindness

$$\kappa_i(C,C) = \frac{\pi_j(C,C) - \pi_j^e(C)}{\pi_j^h(C) - \pi_j^{min}(C)} \tag{2.3}$$

$$= \frac{\frac{2}{3} - \frac{1}{3}}{\frac{2}{3} - 0} \tag{2.4}$$

$$= \frac{1}{2} \tag{2.5}$$

Alternatively, if player i chooses to defect, her action results in negative kindness

$$\kappa_i(D,C) = \frac{\pi_j(C,D) - \pi_j^e(C)}{\pi_j^h(C) - \pi_j^{min}(C)} \tag{2.6}$$

$$= \frac{0 - \frac{1}{3}}{\frac{2}{3} - 0} \tag{2.7}$$

$$= -\frac{1}{2} \tag{2.8}$$

Note that because the kindness function is normalized, its value must lie in the interval $\kappa_i(a_i, a_j^b) \in [-1, \frac{1}{2}]$.

Now consider the following utility function which captures "intention-based" preferences.

$$u_i(a_i, a_i^b) = \pi_i(a_i, a_i^b) + \gamma \left(1 + \kappa_i(a_i, a_i^b)\right) \kappa_j(a_i^b, a_i) \tag{2.9}$$

Each individuals derives utility from her material payoff and also from reciprocating her opponent's kindness. An exogenous weight γ determines how strongly individuals care about reciprocating kindness relative to their material payoff. Note that if player i expects her opponent to exert negative kindness towards her (that is, if $\kappa_j(a_i^b, a_i^{bb}) < 0$), then player i's utility increases if she exerts negative kindness herself. Conversely, if she expects her opponent to exert positive kindness, then player i benefits from doing the same. Hence, individuals with intention-based preferences have a taste for reciprocity: they wish to be kind to those who are kind to them, and unkind to others.

This means that in our prisoners' dilemma example, where player i expects her opponent to cooperate and thus to exert positive kindness, intention-based preferences create an incentive for her to reciprocate with positive kindness as well. Both players expect each other to cooperate, that is, $a_i^b = a_i^{bb} = C$. Hence, player j's expected kindness is given by $\kappa_j(C, C) = \frac{1}{2}$. Player i's utility from cooperating or defecting is then given by:

$$u_i(C, C,) = \frac{2}{3} + \gamma(1 + \frac{1}{2})(\frac{1}{2}) \quad (2.10)$$

$$u_i(D, C) = 1 + \gamma(1 - \frac{1}{2})(\frac{1}{2}) \quad (2.11)$$

Player i will thus find it profitable to cooperate whenever $\gamma \geq \frac{2}{3}$. In this case, if individuals exhibit intention-based preferences we can sustain an equilibrium in which both players cooperate, which Pareto dominates the Nash equilibrium in which both players defect.

2.1.3 Income taxation in the Stiglitz (1982) mechanism

Just as intention-based preferences can reduce individuals' incentives to "defect" in a prisoners' dilemma, as explained in the previous example, the remainder of this paper will explore how they can relax incentive constraints in a model of optimal taxation.

Consider an economy with a continuum of individuals I of mass 1. An allocation $\{(c_k, y_k)_{k \in I}\}$ consists of a bundle of consumption c and output y for every individual. Individuals differ in their productivity types $w_i \in \{w_L.w_H\}$, with $w_L < w_H$. Types are private information and the population share of each type equals one half.

Along the lines of the Stiglitz (1982) model of income taxation, preferences for individuals of some type w_i are given by

$$\pi_i = c - v\left(\frac{y}{w_i}\right) \quad (2.12)$$

I will refer to π_i as the individual's "material payoff." Here, y and c denote individual output and consumption. Material payoff is quasi-linear in consumption and individuals incur a cost of labor $v(\cdot)$ with $v'(\cdot) > 0$ and $v''(\cdot) > 0$.

The social planner offers every individual a set of actions $A = \{a_L, a_H\}$ and specifies an allocation $\{(c_L, y_L), (c_H, y_H)\}$ that maps each action into an individual consumption-output bundle. That is, individuals who choose action a_i will consume c_i and produce y_i.

The planner wishes to redistribute income from highly productive individuals to less productive individuals, while maintaining incentive compatibility and resource constraints. Formally, the planner's objective is

$$\max \quad \alpha\left(\mathrm{c}(\mathrm{a_L}) - \mathrm{v}\left(\frac{\mathrm{y}(\mathrm{a_L})}{\mathrm{w_L}}\right)\right) + (1-\alpha)\left(\mathrm{c}(\mathrm{a_H}) - \mathrm{v}\left(\frac{\mathrm{y}(\mathrm{a_H})}{\mathrm{w_H}}\right)\right) \tag{2.13}$$

$$\text{subject to} \quad \mathrm{c}(\mathrm{a_L}) - \mathrm{v}\left(\frac{\mathrm{y}(\mathrm{a_L})}{\mathrm{w_L}}\right) \geq \mathrm{c}(\mathrm{a_H}) - \mathrm{v}\left(\frac{\mathrm{y}(\mathrm{a_H})}{\mathrm{w_L}}\right) \tag{2.14}$$

$$\mathrm{c}(\mathrm{a_H}) - \mathrm{v}\left(\frac{\mathrm{y}(\mathrm{a_H})}{\mathrm{w_H}}\right) \geq \mathrm{c}(\mathrm{a_L}) - \mathrm{v}\left(\frac{\mathrm{y}(\mathrm{a_L})}{\mathrm{w_H}}\right) \tag{2.15}$$

$$\mathrm{y}(\mathrm{a_L}) - \mathrm{c}(\mathrm{a_L}) + \mathrm{y}(\mathrm{a_H}) - \mathrm{c}(\mathrm{a_H}) \geq 0 \tag{2.16}$$

where $\alpha > 0.5$ is the planner's welfare weight on the material payoff of low types, so that the planner's objective is to redistribute resources towards low types. The two incentive constraints ensure that individuals self-select into the consumption-output bundle that corresponds to their productivity type. Lastly, the resource constraints ensures that total output equals total consumption. The allocation that maximizes the planner's objective subject to these constraints is shown by Stiglitz (1982) and summarized in Definition 2.

Definition 2: *The second-best allocation as shown by Stiglitz (1982) is denoted* $(\mathrm{c_L^{SB}}, \mathrm{y_L^{SB}}, \mathrm{c_H^{SB}}, \mathrm{y_H^{SB}})$ *and is characterized by the following four equations:*

$$\mathrm{v}'\left(\frac{\mathrm{y_H^{SB}}}{\mathrm{w_H}}\right)\frac{1}{\mathrm{w_H}} = 1 \tag{2.17}$$

$$\mathrm{v}'\left(\frac{\mathrm{y_L^{SB}}}{\mathrm{w_L}}\right)\frac{1}{\mathrm{w_L}} = 1 - \left(\mathrm{v}'\left(\frac{\mathrm{y_L^{SB}}}{\mathrm{w_L}}\right)\frac{1}{\mathrm{w_L}} - \mathrm{v}'\left(\frac{\mathrm{y_L^{SB}}}{\mathrm{w_H}}\right)\frac{1}{\mathrm{w_H}}\right) \tag{2.18}$$

$$\mathrm{c_H^{SB}} - \mathrm{v}\left(\frac{\mathrm{y_H^{SB}}}{\mathrm{w_H}}\right) = \mathrm{c_L^{SB}} - \mathrm{v}\left(\frac{\mathrm{y_L^{SB}}}{\mathrm{w_H}}\right) \tag{2.19}$$

$$\mathrm{c_L^{SB}} + \mathrm{c_H^{SB}} = \mathrm{y_L^{SB}} + \mathrm{y_H^{SB}} \tag{2.20}$$

First, the second-best allocation leaves output for high types undistorted, so that the marginal disutility of labor equals the marginal benefit of consumption. Second, the incentive compatibility constraint of low types is binding. This implies that, third, output for low types is distorted downwards. Fourth, the resource constraint is binding.

2.2 Income taxation and intention-based preferences

In this section I propose the optimal mechanism if individuals follow intention-based preferences as a group rule. I augment the Stiglitz (1982) mechanism by offering low type individuals the choice of exerting a higher labor effort. I then construct the optimal allocation $(\mathrm{c_L^*}, \mathrm{y_L^*}, \mathrm{c_H^*}, \mathrm{y_H^*})$ that can be implemented by the mechanism and show that it is Pareto superior to the Stiglitz (1982) second-best allocation $(c_L^{SB}, y_L^{SB}, c_H^{SB}, y_H^{SB})$.

2.2.1 Intention-based preferences in a large economy

To adopt Rabin's two player-model in the context of a large economy, I will define intention-based preferences as a group rule.

As explained in the previous section, Rabin's model in its original form applies to a two-player environment like the prisoners' dilemma, in which an action taken by player i has a direct effect on the material payoff of player j. In a large economy, the environment in which the Stiglitz (1982) income tax model is set, individual actions have no impact on the wellbeing of others. Instead, a person only determines her own material payoff by self-selecting into one of the two possible consumption-output bundles. The material payoff of others is unaffected by this choice. Further, since each person has zero mass, individual actions also have no effect on the public budget. This poses a problem when incorporating Rabin's model, which interprets "kindness" as an individual's effort to increase her opponent's material payoff, into the context of income taxation.

As a solution, I adopt Rabin's model as a so-called group rule, following the work of Harsanyi (1977). Suppose an individual of type w_i chooses some action a_i. In computing her own kindness, the individual does not only consider the causal effect of her own action a_i on the material payoffs of others. Instead, she considers what the effect on the material payoffs of others would be if all individuals of type w_i were to choose action a_i as well.

To formalize this idea, I first define the material payoffs that would arise if all individuals of type w_i were to behave just like individual i.

Definition 3: *For individual* i *with type* $w_i \in \{w_L, w_H\}$*, let*

$$\tilde{\pi}_i(a_i, a_i^b) \tag{2.21}$$

denote i*'s material payoff if all individuals of type* w_i *choose action* a_i *and all individuals of type* $w_j \neq w_i$ *choose action* a_i^b*.*

The kindness function is then adapted as follows:

$$\tilde{\kappa}_i(a_i, a_i^b) = \frac{\tilde{\pi}_j(a_i^b, a_i) - \pi_j^e(a_i^b)}{\pi_j^h(a_i^b) - \pi_j^{min}(a_i^b)} \tag{2.22}$$

where $\pi_j^h(a_i^b)$ and $\pi_j^{min}(a_i^b)$ are, respectively, the highest and lowest material payoff that individuals of type w_j could reach given that they chose action a_i^b. As explained in the previous section, $\pi_j^e(a_i^b)$ is called the equitable payoff, and specified as

$$\pi_j^e(a_i^b) \;=\; \frac{1}{2}\left(\pi_j^h(a_i^b) \;-\; \pi_j^l(a_i^b)\right) \tag{2.23}$$

where $\pi_j^l(a_i^b)$ is the lowest material payoff among all Pareto efficient outcomes possible for individuals of type w_j given that they chose action a_i^b. Again, the kindness function thus always lies within the interval $[-1, \frac{1}{2}]$.

2.2.2 The optimal mechanism

The mechanism offers three actions $A = \{a^H, a^L, a^{L^*}\}$ to each individual. As in the Stiglitz (1982) mechanism, a^H and a^L will be interpreted as a message that the individual has a high or low productivity type, respectively. Additionally, a^{L^*} will be interpreted as a message that the individual has a low productivity type and wishes to exert a higher labor effort $y_L^* > y_L^{SB}$. The mechanism thus augments the Stiglitz (1982) mechanism by offering low type individuals the choice of exerting a higher labor effort.

The outcomes of this mechanism are as follows. Let $(\rho^L, \rho^{L^*}, \rho^H)$ denote the proportion of individuals who play actions a^L, a^{L^*}, and a^H respectively. If $\rho^H \geq \frac{1}{2}$ then everyone who played a^H receives (y_H^*, c_H^*), everyone who played a^{L^*} receives (y_L^*, c_L^*), and everyone who played a^L receives (y_L^{SB}, c_L^{SB}). In this case, if high type individuals truthfully reveal their type, then low type individuals may choose to exert a higher labor effort that results in additional consumption $(c_H^* > c_H^{SB})$ for high types. Otherwise, if $\rho^H < \frac{1}{2}$, then everyone who played a^H receives (y_H^{SB}, c_H^{SB}), and everyone who played a^{L^*} or a^L receives $(0, 0)$.

Intuitively, the idea of this mechanism is to redistribute part of the additional output – which is produced by those low type individuals who play action a^{L^*} and produce y_L^* – towards high type individuals. This way, as will be explained below, exerting a higher labor effort will generates positive kindness for low types because it increases the material payoff for high types. High types can reciprocate positive kindness simply by "revealing" their type, which prevents those who played a^L from receiving zero payoff.

Table 2.2 summarizes the allocations that result from this mechanism if all individuals with the same type choose the same action. This means that Table 2.2 depicts exactly the allocations that individuals take into account for when they follow intention-based preferences as a group rule, as described by Definition 3.

Table 2.2. Allocations that arise if all individuals of the same type play the same action

		High types' action		
		a^L	a^{L^*}	a^H
Low types' action	a^L	$(0,0),(0,0)$	$(0,0),(0,0)$	$(c_L^{SB}, y_L^{SB}),(c_H^{SB}, y_H^{SB})$
	a^{L^*}	$(0,0),(0,0)$	$(0,0),(0,0)$	$(c_L^*, y_L^*),(c_H^*, y_H^*)$
	a^H	$(c_H^{SB}, y_H^{SB}),(c_L^{SB}, y_L^{SB})$	$(c_H^*, y_H^*),(c_L^*, y_L^*)$	$(c_H^*, y_H^*),(c_H^*, y_H^*)$

For example, recall that $\tilde{\pi}_L(a^{L^*}, a^H)$ and $\tilde{\pi}_H(a^H, a^{L^*})$ depict the material payoff for low types and high types if all low types play a^{L^*} and all high types play a^H. In this case, as summarized by Table 2.2, low types receive (c_L^*, y_L^*) and high types receive (c_H^*, y_H^*).

2.2.3 Preferences and equilibrium

Let $s_i \in A$ denote the strategy of an individual of type w_i. Each individual forms first-order and second-order beliefs over the strategies of other players. Let $s_i^b \in A$ denote her first-order belief over the strategy played by individuals of type $w_j \neq w_i$. Further, s_i^{bb} denotes this individual's second-order belief over all other players' beliefs over her strategy.

Utility for an individual of type w_i is given by

$$u_i(s_i, s_i^b, s_i^{bb}) \;=\; \pi_i(s_i, s_i^b) + \gamma\left(1 + \tilde{\kappa}_i(s_i, s_i^b)\right)\tilde{\kappa}_j(s_i^b, s_i^{bb}) \tag{2.24}$$

The following equilibrium definition is adopted from Bierbrauer and Netzer (2016).

Definition: *A Bayes-Nash Fairness Equilibrium (BNFE) is a strategy profile* s^* *such that* $\forall i, j$ *with* $j \neq i$:

$$s_i^* \;\in\; \arg\max_{s_i \in A} u_i(s_i, s_i^b, s_i^{bb}) \tag{2.25}$$

$$s_i^b \;=\; s_j^* \tag{2.26}$$

$$s_i^{bb} \;=\; s_i^* \tag{2.27}$$

In BNFE, first- and second-order beliefs are required to be correct and every individual maximizes their utility given beliefs.

2.2.4 Constructing the optimal allocation

In the following I will construct the optimal allocation $(c^*_L, y^*_L, c^*_H, y^*_H)$ subject to budget and incentive constraints that can be reached as a BNFE by the mechanism described above.

Suppose there exists a BNFE in which all low types play $s^*_L = a^{L^*}$ and all high types play $s^*_H = a^H$. Then for high types the optimal consumption-output bundle is given by

$$y^*_H = y^{SB}_H \tag{2.28}$$
$$c^*_H = c^{SB}_H + \varepsilon \tag{2.29}$$

First, equation 2.28 states that the optimal labor effort for high types is undistorted. This follows the same logic as in Stiglitz (1982). Because intention-based preferences do not affect individual's marginal disutility of labor, it is still optimal to leave high types' labor efforts at their efficient level $y^*_H = y^{SB}_H$. Second, equation 2.29 states that some amount ε of the additional output that is produced by low type individuals who choose a^{L^*} should be redistributed towards high types.

Intuitively, this ensures that high types receive a strictly higher material payoff whenever low types exert extra labor effort. Exerting a higher labor effort will then generate positive kindness for low types, because it increases the material payoff for high types. Clearly, (c^*_H, y^*_H) yields a strictly higher material payoff than (c^{SB}_H, y^{SB}_H), because it elicits the same amount of output but yields higher consumption.

Next, let us specifically evaluate the kindness that a low type individual generates by playing a^{L^*}. If a low type chooses to play a^{L^*} and believes that high types play a^H, then her kindness is given by:

$$\begin{aligned}
\tilde{\kappa}_L(a^{L^*}, a^H) &= \frac{\tilde{\pi}_H(a^H, a^{L^*}) - \pi^e_H(a^H}{\pi^h_H(a^H) - \pi^{min}_H(a^H)} &(2.30)\\
&= \frac{\tilde{\pi}_H(a^H, a^{L^*}) - \frac{1}{2}\left(\pi^h_H(a^H) + \pi^l_H(a^H)\right)}{\pi^h_H(a^H) - \pi^{min}_H(a^H)} &(2.31)\\
&= \frac{\tilde{\pi}_H(a^H, a^{L^*}) - \frac{1}{2}\left(\pi_H(a^H, a^{L^*}) + \pi_H(a^H, w_L)\right)}{\pi_H(a^H, a^{L^*}) - \pi_H(a^H, w_L)} &(2.32)\\
&= \frac{1}{2} &(2.33)
\end{aligned}$$

Thus, exerting a higher labor effort indeed generates positive kindness for low types. If low types play a^{L^*}, high types will then wish to reciprocate by generating positive kindness as well. This creates leeway for the high type's incentive constraint. To see this, suppose high types play a^H and believes that low types will play a^{L^*}. Then, the kindness of high types is given by:

$$\tilde{\kappa}_H(a^H, a^{L^*}) = \frac{\tilde{\pi}_L(a^{L^*}, a^H) - \pi_L^e(a^{L^*})}{\pi_L^h(a^{L^*}) - \pi_L^{min}(a^{L^*})} \tag{2.34}$$

$$= \frac{\tilde{\pi}_L(a^{L^*}, a^H) - \frac{1}{2}\left(\pi_L^h(a^{L^*}) + \pi_L^l(a^{L^*})\right)}{\pi_L^h(a^{L^*}) - \pi_L^{min}(a^{L^*})} \tag{2.35}$$

$$= \frac{\tilde{\pi}_L(a^{L^*}, a^H) - \frac{1}{2}\left(\pi_L(a^{L^*}, a^H) + \pi_L(a^{L^*}, w_L)\right)}{\pi_H(a^{L^*}, a^H) - \pi_L(a^{L^*}, w_L)} \tag{2.36}$$

$$= \frac{1}{2} \tag{2.37}$$

If instead high types were to play a^L or a^{L^*}, while believing that low types play a^{L^*}, then high types would generate negative kindness:

$$\tilde{\kappa}_H(a^{L^*}, a^{L^*}) = \frac{\tilde{\pi}_L(a^{L^*}, a^{L^*}) - \pi_L^e(a^{L^*})}{\pi_L^h(a^{L^*}) - \pi_L^{min}(a^{L^*})} \tag{2.38}$$

$$= \frac{\tilde{\pi}_L(a^{L^*}, a^{L^*}) - \frac{1}{2}\left(\pi_L^h(a^{L^*}) + \pi_L^l(a^{L^*})\right)}{\pi_L^h(a^{L^*}) - \pi_L^{min}(a^{L^*})} \tag{2.39}$$

$$= \frac{\tilde{\pi}_L(a^{L^*}, a^{L^*}) - \frac{1}{2}\left(\pi_L(a^{L^*}, \hat{w}_H) + \pi_L(a^{L^*}, a^{L^*})\right)}{\pi_H(a^{L^*}, \hat{w}_H) - \pi_L(a^{L^*}, a^{L^*})} \tag{2.40}$$

$$= -\frac{1}{2} \tag{2.41}$$

Hence, for high types who believe that low types will play a^{L^*}, "revealing" their type (by playing a^H) generates positive kindness, while "pretending" to be a low type (by playing a^L or a^{L^*}) generates negative kindness. In order to reciprocate the low types' positive kindness, high types will be willing to reveal their type even if this results in a smaller material payoff. Specifically, pretending to be a low type would result in a utility loss of

$$\gamma(1 + \tilde{\kappa}_H(a^H, a^{L^*}))\tilde{\kappa}_L(a^{L^*}, a^H) - \gamma(1 + \tilde{\kappa}_H(a^{L^*}, a^{L^*}))\tilde{\kappa}_L(a^{L^*}, a^H) \tag{2.42}$$

which equals $\frac{\gamma}{2}$.

We can then write a new incentive constraint for high types that includes the utility derived from kindness and will be binding in the optimal allocation:

Lemma 1: *In the optimal allocation* $(c_L^*, y_L^*, c_H^*, y_H^*)$*, the high type's incentive constraint is given by*

$$c_H^* - v\left(\frac{y_H^*}{w_H}\right) \geq c_L^* - v\left(\frac{y_L^*}{w_H}\right) - \frac{\gamma}{2} \tag{2.43}$$

Intention-based preferences thus create slack in the high type's incentive constraint. The new constraint allows the planner to allocate a higher material payoff to low types (compared to the Stiglitz (1982) allocation) that would violate incentive compatibility if individuals did not exhibit intention-based preferences. This way, the planner can also allow smaller distortions of the low type's labor supply.

Equation 2.43 implicitly characterizes the optimal labor supply y_L^* for low types. As a last step, we need to characterize the low type's optimal level of consumption c_L^*. First, note that the public budget constraint is still given by: $c_L^* + c_H^* = y_L^* + y_H^*$. Recall that ε of the additional output produced by low types is redistributed towards high types. Thus, the remaining amount of the additional output will be consumed by low types:

$$c_L^* \;=\; c_L^{SB} + y_L^* - y_L^{SB} - \varepsilon \tag{2.44}$$

Equation 2.44 states that consumption for low types increases by the additional output that they produce minus ε. With ε small, this means that low type individuals essentially move up the isotax curve. Since we know that in the standard second-best allocation y_L^{SB} is distorted downwards

$$v'\left(\frac{y_L^{SB}}{w_L}\right)\frac{1}{w_L} \;<\; 1 \tag{2.45}$$

it is immediately clear that a move up the isotax curve yields an increase in material payoff for low type individuals.

Thus, the optimal allocation $(c_L^*, y_L^*, c_H^*, y_H^*)$ is a Pareto improvement over $(c_L^{SB}, y_L^{SB}, c_H^{SB}, y_H^{SB})$ as both high types and low types receive strictly higher material payoffs and generate positive kindness.

Equations 2.28, 2.29, 2.43 and 2.44 fully characterize the optimal allocation $(c_L^*, y_L^*, c_H^*, y_H^*)$, as summarized by Proposition 1. This allocation is incentive compatible, as explained by Lemma 1. Hence, a Pareto improving BNFE indeed exists where all low type individuals play a^{L^*} and all high type individuals play a^H.

Proposition 1: *The optimal allocation* $(c_L^*, y_L^*, c_H^*, y_H^*)$ *subject to budget and incentive constraints that can be reached in a BNFE of the game described* *above is characterized by*

$$v'\left(\frac{y_H^*}{w_H}\right)\frac{1}{w_H} = 1 \tag{2.46}$$

$$c_H^* - v\left(\frac{y_H^*}{w_H}\right) = c_L^* - v\left(\frac{y_L^*}{w_H}\right) - \frac{\gamma}{2} \tag{2.47}$$

$$c_H^* = c_H^{SB} + \varepsilon \tag{2.48}$$

$$c_L^* = c_L^{SB} + y_L^* - y_L^{SB} - \varepsilon \tag{2.49}$$

where $\varepsilon > 0$ is small, and constitutes a Pareto improvement over the standard second-best allocation.

Note that $\varepsilon > 0$ should be as small as possible to maximize welfare in the optimal allocation. By optimality of the standard second-best allocation we know that the welfare gain from marginally increasing y_L^* exactly equals the welfare loss from compensating the high type in order to restore incentive compatibility. This equality is captured by equation 2.18. This is equivalent to saying that the welfare gain from marginally increasing c_H^* and creating slack in the high type's incentive constraint is equal to the welfare loss from increasing y_L^* to restore budget feasibility. In the optimal allocation, with $y_L^{SB} < y_L^*$, quasi-linearity and single-crossing imply that:

$$1 - v'\left(\frac{y_L^*}{w_L}\right)\frac{1}{w_L} - \left(v'\left(\frac{y_L^*}{w_L}\right)\frac{1}{w_L} - v'\left(\frac{y_L^*}{w_H}\right)\frac{1}{w_H}\right) < 0 \tag{2.50}$$

Thus, any increase in c_H^* and an associated increase in y_L^* to maintain binding incentive and budget constraints would lead to a strict welfare loss. Thus, in the optimal allocation ε should be as small as possible.

Graphically, Figure 2.1 depicts the optimal allocation $(c_L^*, y_L^*, c_H^*, y_H^*)$, in the BNFE in which all high types play a^H and all low types play a^{L^*}. The consumption for high types is increased by ε. The optimal allocation for low types is bounded from above by the budget constraint (depicted by the dashed 45° isotax curve) and the high type's incentive constraint (depicted by the dashed red line). Since material payoff for the low type is increasing as he moves up the isotax curve and decreasing as he moves up the high type's indifference curve, the optimal allocation lies at the intersection of the two.

Robustness against low prevalence of intention-based preferences

As an extension, I show that the BNFE from Proposition 1 exists even if a (sufficiently small) share of the population does not exhibit intention-based preferences.

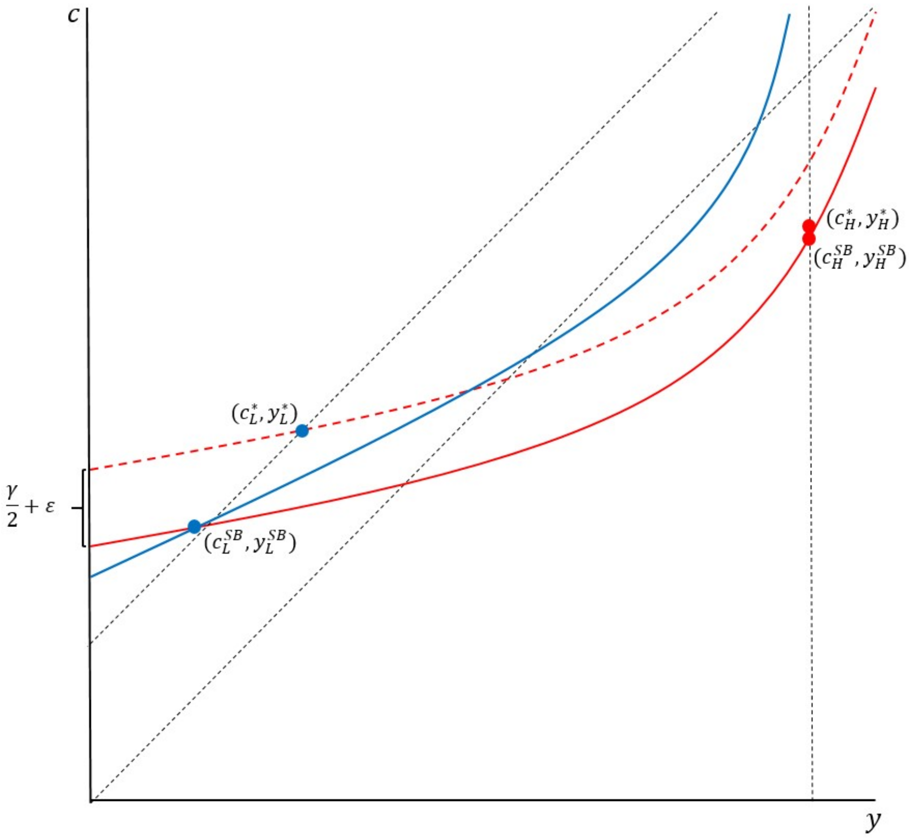

Figure 2.1

Let f_γ denote the proportion of individuals who exhibit intention-based preferences. Then, a share $(1 - f_\gamma)$ of individuals is "selfish" in the sense that they only care about their material payoffs, which can be expressed by setting $\gamma = 0$ in their utility functions. I will maintain the assumption that the population shares of high and low productivity types are equal, and also assume that the share of selfish individuals is the same for each productivity type.

I further assume that individuals with intention-based preferences will take into account the kindness of all other individuals, even those who are selfish. The kindness of high type individuals as a group will be given by the sum of individual kindness weighted by the population shares. Lastly, I will modify the mechanism described in the previous section by implementing $(c_L^*, y_L^*, c_H^*, y_H^*)$ even when $\rho^H < \frac{1}{2}$. In this case, consumption for low types c_L^* will be lowered in order to maintain a balanced budget.

It is immediately clear that in the BNFE described by Propisition 1, selfish high type individuals would pretend to be a low type and play a^{L^*}, since this yields the highest material payoff. This has two implications which may limit the existence of a BNFE as described by Proposition 1.

First, selfish high type individuals pretending to be a low type will be perceived as unkind by individuals with intention-based preferences. Thus, if too many high type individuals are selfish, then high types as a group will be perceived as unkind and low types with intention-based preferences will want to act unkindly, thus making it impossible to sustain a welfare improving BNFE. Here, kindness of high types as a group is given by:

$$\begin{aligned} & f_\gamma \tilde{\kappa}_H(a^H, a^{L^*}) + (1 - f_\gamma)\tilde{\kappa}_H(a^{L^*}, a^{L^*}) && (2.51)\\ = \ & f_\gamma \frac{1}{2} + (1 - f_\gamma)(-\frac{1}{2}) && (2.52)\\ = \ & f_\gamma - \frac{1}{2} && (2.53) \end{aligned}$$

Thus, for high types' kindness to be perceived as positive on the aggregate we need $f_\gamma > \frac{1}{2}$.

Second, selfish high types who play a^{L^*} put pressure on the resource constraint. If only a share f_γ of high types produces more output than they consume, then less resources are available to be redistributed to low type individuals in equilibrium. This, too, makes it difficult to sustain a Pareto improving BNFE if sufficiently many high types are selfish. The new resource constraint is then given by:

$$y_L^* - c_L^* + f_\gamma(y_H^* - c_H^*) + (1 - f_\gamma)(y_L^* - c_L^*) \ = \ 0 \qquad (2.54)$$

Hence of every unit of consumption that the planner intents to extract from high types, he can only redistribute a share $\frac{f_\gamma}{2-f_\gamma}$ to low types. The rest is lost because of the behavioral response of those high type individuals that have selfish preferences.

Graphically, as f_γ decreases, (c_L^*, y_L^*) moves upwards along the indifference curve of high types in order to maintain budget feasibility, which yields a decrease in material payoff for low types. Let $\underline{f}$ denote the smallest value for f_γ such that low types still prefer (c_L^*, y_L^*) over (c_L^{SB}, y_L^{SB}). If $f_\gamma < \underline{f}$ then $(c_L^*, y_L^*, c_H^*, y_H^*)$ is no longer sustained by the BNFE described in Proposition 1 because low types are strictly worse off than in the standard second-best allocation.

Hence, in order to sustain a Pareto improving BNFE as described by Proposition 1, we need that

$$f_\gamma \ > \ \max\{\frac{1}{2}, \underline{f}\} \qquad (2.55)$$

2.3 Conclusion

The paper demonstrates that in the presence of intention-based preferences, we can implement an allocation as the unique BNFE that Pareto dominates the Stiglitz (1982) allocation. Allowing the "poor" to exert a higher labor effort generates positive kindness which is reciprocated by the "rich" and creates slack in their incentive constraints. The interpretation is that the rich do not mind helping the poor if they "do their part" by working harder. Importantly, I show that there exists a nonwelfarist principle which facilitates a Pareto superior allocation. Because intention-based preferences constitute a notion of procedural justice, the implementation of this allocation requires altering individuals' choice sets. It is thus an allocation that cannot be characterized by the maximization of any weighted social welfare function.

References

Bassetto, Marco, and Christopher Phelan. 2008. “Tax Riots.” *Review of Economic Studies* 75 (3): 649–69. [19]

Bierbrauer, Felix, and Nick Netzer. 2016. “Mechanism design and intentions.” *Journal of Economic Theory* 163 (C): 557–603. [20, 29]

Coate, Stephen, and Michael Conlin. 2004. “A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence.” *American Economic Review* 94 (5): 1476–504. [21]

Feddersen, Timothy, and Alvaro Sandroni. 2006. “A Theory of Participation in Elections.” *American Economic Review* 96 (4): 1271–82. [21]

Harsanyi, John C. 1977. “Rule utilitarianism and decision theory.” *Erkenntnis* 11 (1): 25–53. [21, 27]

Harsanyi, John C. 1980. “Rule utilitarianism, rights, obligations and the theory of rational behavior.” *Theory and Decision* 12 (2): 115–33. [21]

Rabin, Matthew. 1993. “Incorporating Fairness into Game Theory and Economics.” *American Economic Review* 83 (5): 1281–302. [20, 22]

Saez, Emmanuel, and Stefanie Stantcheva. 2016. “Generalized Social Marginal Welfare Weights for Optimal Tax Theory.” *American Economic Review* 106 (1): 24–45. [19–21]

Slemrod, Joel B. 2000. *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*. Harvard University Press. [19]

Stiglitz, Joseph E. 1982. “Self-selection and Pareto efficient taxation.” *Journal of Public Economics* 17 (2): 213–40. [20–22, 25–28, 30, 32, 36]

Weinzierl, Matthew. 2014. “The promise of positive optimal taxation: normative diversity and a role for equal sacrifice.” *Journal of Public Economics* 118: 128–42. [21]

Weinzierl, Matthew. 2017. “A Welfarist Role for Nonwelfarist Rules: An Example with Envy.” *Harvard Business School Working Paper No. 17-021*, [21]

Chapter 3

Demand for think tanks in the presence of (non-)partisan experts

3.1 Introduction

In the United States, 1,839 "think tanks" operated in 2014. McGann (2016) posits that these think tanks take on "critical roles in the policymaking process by offering original research and analysis, [and] providing policy advice" which enable policy makers to make informed decisions. Further, many think tanks act as advocacy groups and "possess extremely strong ideologies or focus on persuading policymakers and the public on short-term, specific policy debates." Examples of such advocacy-oriented think tanks include the Heritage Foundation and the Center for American Progress, respectively on the right and left of the political spectrum. In the United States, 36% of think tanks self-identify as partisan – representing either conservative, libertarian, or progressive positions – and "view their role in the policymaking process as winning the war of ideas rather than disinterestedly searching for the best policies." (McGann, 2016). Hence, as an industry, these think tanks produce partisan research to inform and influence policy makers.

One way in which think tanks influence policy makers is by providing information. As McGann observes, "U.S. politicians and bureaucrats have increasingly turned to think tanks to provide research and systematic analysis that are reliable, policy relevant, and, above all, useful." In the context of public finance, relevant research concerns in particular the estimation of economic parameters that are crucial for the optimal design of fiscal policy. For example, policy makers require information about the size of the deadweight loss caused by taxation, or the extent to which redistribution shrinks the size of the "economic pie." Prominent examples of such parameters are the elasticity of taxable income and the efficiency of government. Policy makers rely on the correct estimation of these parameters for the optimal choice of progressivity of the tax system and size of the government. McGann points out that the relationship between think tanks and policy makers "mimics that of a market – a market of knowledge. The demand for think tanks' policy research is driven by the legislative demand for relevant information."

This existing "market of knowledge" as described by McGann poses a puzzle, however, when considering that a large academic literature produces research on the same topics (see Saez, Slemrod, and Giertz (2012) for a recent overview). The question arises, how can demand for think tanks co-exist with the vast amount of information that is made publicly available by academic researchers? This paper aims to answer this question and ventures that demand for think tanks can only exist if academic research is itself partisan.

The paper proposes a model of Bayesian voters who differ in their pre-tax incomes and are uncertain about the deadweight loss caused by taxation. For example, they might be uncertain about the elasticity of taxable income or the government's administrative cost. Individuals form beliefs over the inefficiency of taxation and vote over a linear income tax rate. The government then uses tax revenue to provide a public good. Voters benefit from more precise information about the "state of the world," that is, the true efficiency of the tax system. If individuals believe that the deadweight loss caused by taxation is small, they will (ceteris paribus) prefer a higher tax rate. An academic expert observes the true state of the world and communicates his information via a cheap talk message to the electorate. Dependent on how credible the expert's message is, individuals may choose to pay for a think tank. Think tanks are assumed to have access to the same information as an academic expert, but they have a partisan bias. That is, think tanks are modeled as strategic communicators that represent the interests of a certain part of the electorate, rather than trying to truthfully inform all individuals about the state of the world. I show that in the presence of a non-partisan academic expert, there exists no demand for think tanks. Indeed, demand for think tanks exists if and only if there is a strictly positive probability that the academic expert is partisan himself.

The paper builds on previous work on electoral competition over linear tax rates, such as Roberts (1977), and adds to it by considering voters who are uncertain about the efficiency of the tax system. Information about this efficiency is conveyed in a game of cheap talk by Crawford and Sobel (1982). Previous

1. In the United States, think tanks like the Cato Institute, the Heritage Foundation, and the American Enterprise Institute, for example, respectively receive 83%, 59%, and 58% of their income from individual contributions (McGann, 2016).

research has studied the effect of cheap talk lobbying on electoral competition, for example by Austen-Smith and Wright (1992) who study strategic communication between lobbyists and elected legislators. However, to the best of my knowledge, no previous work has studied lobbying as strategic communication with the electorate. This paper is most closely related to Roemer (1994) who analyzes the origin of party ideology when voters are uncertain about the efficiency of redistribution. Roemer, however, considers voters who are not Bayesian and would be influenced even by non-credible information. The question why policy makers rely on information provided by partisan think tanks, as opposed to non-partisan academics, also relates to work by Dewatripont, Jewitt, and Tirole (1999), who study why decision makers prefer to rely on advocates of special interests rather than an unbiased representative. In their setting, a principal (for example a judge) hires agents (lawyers) to investigate two competing causes by collecting evidence that is ex-post verifiable. Because the agents' rewards are conditional on the principal's decision and not on the information collected, there exists an incentive problem, as it is costly to incentivize a non-partisan agent to investigate two causes if finding contradictory evidence (that is, evidence in support of both causes) results in zero rewards. In their setting it is thus desirable for the principal to hire two partisan agents who are rewarded only if the principal's decision favors their cause. Lastly, political scientists have also studied demand for think tanks. For example, empirical work by Bertelli and Wenger (2009) suggests that legislative debate as modeled by Austen-Smith drives demand for strategic information and creates a market opportunity for political entrepreneurs to create think tanks.

In addition to contributing to the literature on the applied theory of public finance, this paper also hopes to pose an interesting question for economic researchers to reflect on.

3.2 The model

There is an economy with n citizens who differ in their skill types $\omega_1, ..., \omega_n$ with $\omega_i < \omega_j \Leftrightarrow i < j$. An individual of type ω_i has preferences given by

$$u(c, G, y, \omega_i) \;=\; c + G - \frac{\alpha}{1+\alpha}\left(\frac{y}{\omega_i}\right)^{\frac{1+\alpha}{\alpha}} \tag{3.1}$$

where c is consumption of a private good, G is consumption of a public good, and y is pre-tax income. Individuals choose their optimal pre-tax income

$$y(\tau, \omega_i) \;=\; \arg\max_{y} \; u(c, G, y, \omega_i) \text{ subject to } c \;=\; (1-\tau)y \tag{3.2}$$

where $\tau \in [0, \overline{\tau}]$ is a linear tax rate.

3.2.1 Taxes and public good provision

The government collects tax revenue and provides a public good $G(\tau, \theta)$, which depends on the tax rate τ as well as on the efficiency cost of taxation $\theta \in \{\theta^L, \theta^H\}$, $\theta^L < \theta^H$. Individuals cannot observe the true value of θ. Instead, they only know its prior distribution, namely that $\theta = \theta^H$ with probability p^H, and $\theta = \theta^L$ with probability $(1 - p^H)$.

In the following subsections I outline two alternative ways in which uncertainty over inefficiency in the public sector may occur and thus how the parameter θ may be interpreted. First, I will consider uncertainty over the government's productivity of providing the public good G. Second, I will consider uncertainty over the extent of incentive problems of taxation, that is, uncertainty over the elasticity of taxable income. These two cases should be understood as alternative specifications for this model. Either one could serve as an interpretation of the underlying uncertainty represented by the "state of the world" θ.

3.2.1.1 Alternative I: Uncertainty over the government's productivity

First, suppose that individuals are uncertain over the government's productivity of public good provision. As the government collects tax revenue and produces a public good G, it incurs administrative costs that are a deadweight loss to society. Specifically, of all the dollars of tax revenue collected, only a share θ will be invested in public good provision. The government's production function is then given by:

$$G(\tau, \theta) = \frac{\theta}{n} \sum_{i=1}^{n} \tau y(\tau, \omega_i) \tag{3.3}$$

where $\theta \in \{\theta^L, \theta^H\}$ is the government's productivity with respect to public good provision. Thus, government productivity is high with probability p^H and low with probability $1 - p^H$. Pre-tax incomes do not depend on θ and are given by

$$y(\tau, \omega_i) = (1 - \tau)^{\alpha} \omega_i, \tag{3.4}$$

hence the elasticity of taxable income is known and equal to

$$\varepsilon_i = \frac{(1 - \tau)}{y(\tau, \omega_i)} \frac{\partial y(\tau, \omega_i)}{\partial (1 - \tau)} = \alpha. \tag{3.5}$$

3.2.1.2 Alternative II: Uncertainty over the elasticity of taxable income

Alternatively, suppose that individuals are uncertain over the aggregate elasticity of taxable income. Any increase in the tax rate reduces every individual's incentive to work and thus lowers the total output in the economy. In addition, the richest individuals may find it profitable to engage in tax avoidance, which has a negative effect on the total amount of tax revenue collected. But the size of this effect may be difficult to estimate for all other individuals who do not have access to instruments of tax avoidance. Thus, individuals are uncertain about the size of the deadweight loss that is caused by taxation and the financing of the public good G. Specifically, let a small number (n − k) of the richest individuals have an elasticity of taxable income θ that is unknown to the rest of the population. Preferences then differ for these individuals as follows:

$$u(c, G, y, \omega_i) \begin{cases} c + G - \frac{\alpha}{1+\alpha}\left(\frac{y}{\omega_i}\right)^{\frac{1+\alpha}{\alpha}} & \text{, for } i = 1, ..., n-k-1 \\ c + G - \frac{\theta}{1+\theta}\left(\frac{y}{\omega_i}\right)^{\frac{1+\theta}{\theta}} & \text{, for } i = n-k, ..., n \end{cases} \tag{3.6}$$

Pre-tax incomes are then given by

$$y(\tau, \omega_i) = \begin{cases} (1-\tau)^{\alpha}\omega_i & \text{, for } i = 1, ..., n-k-1 \\ (1-\tau)^{\theta}\omega_i & \text{, for } i = n-k, ..., n \end{cases} \tag{3.7}$$

where it is assumed that $\frac{\omega_{n-k}}{\omega_{n-k-1}} > (1-\overline{\tau})^{\alpha-\theta_H}$, so that pre-tax incomes are strictly increasing in the skill type. Individual elasticities of taxable income are equal to

$$\varepsilon_i = \frac{(1-\tau)}{y(\tau, \omega_i)} \frac{\partial y(\tau, \omega_i)}{\partial(1-\tau)} = \begin{cases} \alpha & \text{, for } i = 1, ..., n-k-1 \\ \theta & \text{, for } i = n-k, ..., n \end{cases} \tag{3.8}$$

and the aggregate elasticity of taxable income is then given by

$$\varepsilon = \frac{(1-\tau)}{\sum_{i=1}^{n} y(\tau, \omega_i)} \frac{\partial \sum_{i=1}^{n} y(\tau, \omega_i)}{\partial(1-\tau)} \tag{3.9}$$

$$= \frac{(1-\tau)}{\sum_{i=1}^{n} y(\tau, \omega_i)} \sum_{i=1}^{n} \frac{\partial y(\tau, \omega_i)}{\partial(1-\tau)} \tag{3.10}$$

$$= \frac{1}{\sum_{i=1}^{n} y(\tau, \omega_i)} \sum_{i=1}^{n} \varepsilon_i y(\tau, \omega_i) \tag{3.11}$$

$$= \frac{1}{\sum_{i=1}^{n} y(\tau, \omega_i)} \left(\alpha \sum_{i=1}^{n-k-1} y(\tau, \omega_i) + \theta \sum_{i=n-k}^{n} y(\tau, \omega_i) \right) \tag{3.12}$$

$$\tag{3.13}$$

Clearly, individuals are uncertain over the aggregate elasticity of taxable income, because it depends on the true value of θ. Finally, the government's production function then equals

$$G(\tau, \theta) = \frac{1}{n} \sum_{i=1}^{n} \tau y(\tau, \omega_i) \tag{3.14}$$

which depends on θ because $y(\tau, \omega_i)$ depends on θ for $i = (n - k), ..., n$. In this specification there is no administrative cost of the government, hence every dollar of tax revenue collected will be invested in public good provision.

3.2.2 Expert messages and individual beliefs

The true value of θ is observed only by an expert. Consider for example an academic institution that conducts research on the elasticity of taxable income or the government's productivity. The academic expert can costlessly observe θ and communicate the value via a cheap talk message $\sigma_E \in \{\sigma^L, \sigma^H\}$ indicating that the deadweight loss caused by taxation is low or high, respectively. This message is observed by all individuals. The expert's preferences will be described in detail below. It will be clear that the expert's preferences are relevant for his ability to credibly transmit information about θ. Given the expert's message σ_E, all individuals form posterior beliefs over the value of θ. This posterior belief is defined as follows:

Definition 1: *Let* b *be the individual belief over the probability that the government's efficiency is high, that is,*

$$b = \Pr[\theta = \theta^H \mid \sigma_E] \tag{3.15}$$

The posterior belief b will be the same for all individuals, because all individuals have the same prior belief and observe the same message σ_E sent by the expert. An individual's (expected) indirect utility is then denoted

$$E_\theta[V(\tau, \theta, \omega_i) \mid b] = (1 - \tau)y(\tau, \omega_i) + E_\theta[G(\tau, \theta) \mid b] - \frac{\alpha}{1 + \alpha}\left(\frac{y(\tau, \omega_i)}{\omega_i}\right)^{\frac{1+\alpha}{\alpha}} \tag{3.16}$$

where only the amount of public good provision is uncertain and its expectation is dependent on the posterior b.

2. In this alternative specification one can argue that $(n - k)$ individuals would know the true value of θ with certainty, because it appears in their utility function. This is not a cause for concern, however, if $(n - k)$ is only a small number of people (say, the top 1% of the income distribution). This group does not include the median voter and, as will become clear in the analysis of the model below, is unable to credibly transmit information about the true value of θ to anybody else.

3.2.3 Voting over taxes

All individuals vote over the linear tax rate τ. The tax rate that is a Condorcet winner in this election will be implemented. There are many voting processes that implement a Condorcet winner, most famously the model by Black (1948) in which individuals engage in pairwise majority voting, or the Downs (1957) model of political competition in a two-party system. Rather than choosing a specific voting process in this paper, I will outline that there exists a Condorcet winner which can be implemented by any of these well known models of political competition.

Lemma 1: *The median voter's preferred tax rate*

$$\tau(\mathrm{b}, \omega_\mathrm{M}) = \arg\max_{\tau \in [0,\overline{\tau}]} \mathrm{E}_\theta[\mathrm{V}(\tau, \theta, \omega_\mathrm{M}) \mid \mathrm{b}] \tag{3.17}$$

where M *is the median of* $1, ..., \mathrm{n}$*, is a Condorcet winner among all tax rates* $[0, \overline{\tau}]$.

<u>Proof:</u> Note first that using the Envelope theorem we have

$$\frac{\partial}{\partial \tau}\mathrm{E}_\theta[\mathrm{V}(\tau, \theta, \omega_\mathrm{i}) \mid \mathrm{b}] = -\mathrm{y}(\tau, \omega_\mathrm{i}) + \frac{\partial}{\partial \tau}\mathrm{E}_\theta[\mathrm{G}(\tau, \theta) \mid \mathrm{b}] \tag{3.18}$$

$$\text{and} \quad \frac{\partial^2}{\partial \tau \partial \omega_\mathrm{i}}\mathrm{E}_\theta[\mathrm{V}(\tau, \theta, \omega_\mathrm{i}) \mid \mathrm{b}] = -\frac{\partial}{\partial \omega_\mathrm{i}}\mathrm{y}(\tau, \omega_\mathrm{i}) < 0 \tag{3.19}$$

This implies that the following single-crossing property holds, which is introduced in Gans and Smart (1996): If an individual of type ω_i prefers a low tax rate τ over $\tau' > \tau$, then all individuals who are richer than individual i will also prefer τ over τ'. Formally:

$$\begin{aligned} \text{For } \tau' > \tau, \quad & \mathrm{E}_\theta[\mathrm{V}(\tau, \theta, \omega_\mathrm{i}) \mid \mathrm{b}] > \mathrm{E}_\theta[\mathrm{V}(\tau', \theta, \omega_\mathrm{i}) \mid \mathrm{b}] \\ \Rightarrow \quad & \mathrm{E}_\theta[\mathrm{V}(\tau, \theta, \omega_\mathrm{j}) \mid \mathrm{b}] > \mathrm{E}_\theta[\mathrm{V}(\tau', \theta, \omega_\mathrm{j}) \mid \mathrm{b}] \quad \forall \mathrm{j} > \mathrm{i} \end{aligned} \tag{3.20}$$

Conversely, if individual i prefers τ over a lower tax rate $\tau'' < \tau$, then all individuals who are poorer than individual i will also prefer τ over τ'':

$$\begin{aligned} \text{For } \tau'' < \tau, \quad & \mathrm{E}_\theta[\mathrm{V}(\tau, \theta, \omega_\mathrm{i}) \mid \mathrm{b}] > \mathrm{E}_\theta[\mathrm{V}(\tau'', \theta, \omega_\mathrm{i}) \mid \mathrm{b}] \\ \Rightarrow \quad & \mathrm{E}_\theta[\mathrm{V}(\tau, \theta, \omega_\mathrm{j}) \mid \mathrm{b}] > \mathrm{E}_\theta[\mathrm{V}(\tau'', \theta, \omega_\mathrm{j}) \mid \mathrm{b}] \quad \forall \mathrm{j} < \mathrm{i} \end{aligned} \tag{3.21}$$

Consequently this property implies that half the population (all individuals with an income smaller than the median voter's) prefers $\tau(\mathrm{b}, \omega_\mathrm{M})$ over any lower tax rate. Similarly, the other half of the population (all individuals with an income greater than the median voter's) prefers $\tau(\mathrm{b}, \omega_\mathrm{M})$ over any higher tax rate. Thus, $\tau(\mathrm{b}, \omega_\mathrm{M})$ is a Condorcet winner among all tax rates.

If the deadweight loss caused by taxation is expected to be lower – that is, as b increases – the median voter will ceteris paribus prefer a higher tax rate: $\frac{\partial\tau(b,\omega_M)}{\partial b} \geq 0$. Thus, information about the true value of θ has political relevance because it influences the tax rate that is implemented in the election.

3.2.4 Timing and equilibrium

The timing of the model is as follows. First, nature determines the value of $\theta \in \{\theta^L, \theta^H\}$, which is only observed by the expert. Second, the expert sends a message $\sigma_E \in \{\sigma^L, \sigma^H\}$, which is observed by everybody, to communicate the value of θ. Given the expert's message, individuals update their belief b over θ. Third, the tax rate that is a Condorcet winner given individuals' belief b is implemented. Fourth, individuals choose their pre-tax incomes $y(\tau, \omega)$ and the government collects tax revenue to provide the public good $G(\tau, \theta)$.

The equilibrium concept is that of Weak Perfect Bayesian Equilibrium, in which the following conditions hold.

- Each individual chooses their pre-tax income $y(\tau, \omega_i)$ optimally given their belief b over the efficiency of taxation.
- The tax rate τ is the Condorcet winner among all tax rates that are individually optimal given belief b.
- Individuals update their belief $b = \Pr[\theta = \theta^H \mid \sigma_E]$ according to Bayes' rule after observing the expert's message σ_E.
- The expert sends the message σ_E that maximizes his payoff in anticipation of individual belief formation, the Condorcet winning tax rate, and individual incomes.

3.3 Non-partisan expert

As a benchmark case, I will first consider an expert who is non-partisan. Consider an academic expert who has no preferences over taxes and who's only objective is to truthfully announce the value of θ. This means that all individuals will be perfectly informed about the efficiency of taxation by observing the expert's message. Since everyone will know the value of θ with certainty, their belief will be given by $b = 0$ if $\sigma_E = \sigma^L$, or $b = 1$ if $\sigma_E = \sigma^H$. By the median voter result described above, the equilibrium tax rate will be given by:

$$\tau(b, \omega_M) \text{ with } b = \begin{cases} 0 \text{ if } \theta = \theta^L \\ 1 \text{ if } \theta = \theta^H \end{cases} \tag{3.22}$$

Demand for a think tank

Suppose that individuals are given the option to pool their resources and build a think tank that acts as an additional expert. That is, a think tank would be able to observe the true value of θ and send a message $\sigma_{\mathrm{TT}} \in \{\sigma^{\mathrm{L}}, \sigma^{\mathrm{H}}\}$ to inform the electorate about the government's efficiency. It is clear that there is no demand for a think tank in the presence of a non-partisan expert, because everyone is already perfectly informed in equilibrium. This means that if an academic expert who is non-partisan truthfully reveals the efficiency of the government, then no individual should be willing to spend resources on a think tank.

3.4 Partisan expert

For the remainder of this paper I will consider the case in which the expert is partisan. That is, the expert's objective is to maximize the utility of some individual of type ω_{E}. In the following I will refer to ω_{E} as the expert's "bias," because the expert's preferences deviate from the preferences of the median voter. If the expert is partisan, or "biased," his message becomes strategic communication in a cheap talk game with all individuals. The expert anticipates how his message affects individual beliefs and the resulting equilibrium tax rate, and sends the message that maximizes the equilibrium utility of individuals of type ω_{E}.

In this section I will characterize the strategic communication game and the equilibrium tax rates. I first assume that the expert's preferences are publicly known. In this case, individuals understand the expert's incentives and can properly interpret his message to update their beliefs. Afterwards, I repeat the analysis for the case in which individuals are uncertain about the expert's preferences.

3.4.1 Known bias

Suppose the expert's preferences are known to all individuals. Building on the analysis from Crawford and Sobel (1982), it becomes clear that the communication game between the expert and all individuals has two possible equilibria: one in which the true value of θ is fully revealed by the expert, and one in which the expert "babbles" and discloses no information about θ. Which equilibrium occurs depends on the expert's bias. The expert understands that revealing his information about the government's efficiency will influence everybody's preferences over taxes and thus affect the equilibrium tax rate. The expert will be willing to reveal the government's efficiency if and only if it is beneficial for him to influence the equilibrium tax rate in this direction. In other words, the fully informative equilibrium occurs if and only if the expert's and the median voter's preferences are sufficiently aligned. To show this, I introduce the following definition:

Definition 2: *Let $\underline{\omega}$ and $\overline{\omega}$, with $\omega_1 < \underline{\omega} < \overline{\omega} < \omega_n$, be given by:*

$$V(\tau(0,\omega_M),\theta^L,\underline{\omega}) \;=\; V(\tau(1,\omega_M),\theta^L,\underline{\omega}) \tag{3.23}$$

$$V(\tau(1,\omega_M),\theta^H,\overline{\omega}) \;=\; V(\tau(0,\omega_M),\theta^H,\overline{\omega}) \tag{3.24}$$

Recall that $\tau(0,\omega_M)$ and $\tau(1,\omega_M)$ are the only two tax rates that can result in equilibrium if all individuals are perfectly informed, as stated by equation (3.22). Equation (3.23) then defines a skill type $\underline{\omega}$ such that any individual with a type $\omega < \underline{\omega}$ will always want the median voter to implement $\tau(1,\omega_M)$, the higher of the two possible tax rates. Conversely, equation (3.24) defines a skill type $\overline{\omega}$ such that any individual with a type $\omega > \underline{\omega}$ will always want the median voter to implement $\tau(0,\omega_M)$, the lower of the possible tax rates.

This implies that any expert with a bias $\omega_E < \underline{\omega}$ has no incentive to reveal that the efficiency of taxation is low. Such an expert always prefers high taxes, no matter what the true state of the world is. Hence, if individuals were to believe the expert's message, then he should always send the pro-redistribution message σ^H. But since individuals understand his conflict of interest they will never believe any message that the expert sends. On the other hand, any expert with bias $\omega_E > \overline{\omega}$ will never find it profitable to reveal that the government's efficiency is high, because he always prefers low taxes regardless of the state of the world. Thus, whenever the expert's bias is $\omega_E \notin [\underline{\omega},\overline{\omega}]$, the communication game between the expert and all individuals results in a "babbling" equilibrium which is completely uninformative.

The only case in which the expert has an incentive to truthfully reveal the value of θ is when $\omega_E \in [\underline{\omega},\overline{\omega}]$. From Definition 2 it is clear that in this case it holds that:

$$V(\tau(0,\omega_M),\theta^L,\omega_E) \;>\; V(\tau(1,\omega_M),\theta^L,\omega_E) \tag{3.25}$$

and

$$V(\tau(1,\omega_M),\theta^H,\omega_E) \;>\; V(\tau(0,\omega_M),\theta^H,\omega_E) \tag{3.26}$$

That is, it is always in the expert's interest to send a truthful message about the government's efficiency, because the expert's objective is maximized whenever the median voter is fully informed. Because the expert's bias is known, individuals understand the expert's incentives to lie or to speak the truth. Thus, they believe the expert's message if and only if $\omega_E \in [\underline{\omega},\overline{\omega}]$, that is, whenever the expert's and median voter's preferences are sufficiently aligned. In this case, they will update their belief to $b = 0$ if $\sigma_E = \sigma^L$, or $b = 1$ if $\sigma_E = \sigma^H$.

Demand for a think tank

Now, consider again the opportunity for individuals to create a think tank that acts just like an additional expert. A think tank can observe the true value of θ and sends a message $\sigma_{\mathrm{TT}} \in \{\sigma^{\mathrm{L}}, \sigma^{\mathrm{H}}\}$ to inform the electorate about the government's efficiency. Analogously to the expert, a think tank will have a known bias ω_{TT}, that is, it represents the interests of some individual of type ω_{TT} which is known to everybody. In the following passage, I characterize individuals' demand for such a think tank. I use the term "demand" to describe an individual's willingness to pay for a think tank at time zero, that is, before the expert sends his message. In other words, demand for a think tank equals the difference between an individual's expected equilibrium utility in absence of a think tank and in the presence of a think tank.

It is clear that if $\omega_{\mathrm{E}} \in [\underline{\omega}, \overline{\omega}]$, then there is no demand for a think tank. In this case, the expert's "bias" is sufficiently small so that he always truthfully reveals the government's efficiency. Thus, even though the expert is partisan, he acts exactly like a non-partisan expert. Just like in our benchmark case with a non-partisan expert, everyone is fully informed and has no demand for a think tank.

However, if $\omega_{\mathrm{E}} \notin [\underline{\omega}, \overline{\omega}]$, then demand for a think tank exists. In this case the expert is unable to communicate any information about θ whatsoever. If a think tank can effectively convey information about the government's efficiency, there will be a positive number of individuals who are willing to pay for the think tank. Note that for a think tank to credibly communicate any information it must hold that $\omega_{\mathrm{TT}} \in [\underline{\omega}, \overline{\omega}]$. That is, there only exists demand for a think tank with a bias that is small enough for it to fully inform voters about the value of θ. This would be a think tank that acts like a non-partisan expert and always sends a truthful message about the government's efficiency.

Let D^{KB} denote an individual's demand for a think tank with known bias $\omega_{\mathrm{TT}} \in [\underline{\omega}, \overline{\omega}]$, that is, a think tank that truthfully reveals the government's efficiency. As mentioned above, individual demand equals the difference between expected equilibrium utility in the presence of a think tank (when everyone is fully informed) and in absence of a think tank (when everyone has no information

3. I will only characterize demand for a think tank and not model the supply side. Creating a think tank clearly involves costs that could outweigh aggregate willingness to pay. Further, the resources that individuals can mobilize might be smaller than their aggregate willingness to pay due to free-riders and other collective action problems. Providing a model for the creation of think tanks goes beyond the scope of this paper and I will solely characterize the demand side of the market for think tanks.

about θ other than their prior). Then individual demand is given by

$$\begin{aligned} \mathrm{D}^{\mathrm{KB}}(\omega_{\mathrm{i}}) \;=\; & \mathrm{p}^{\mathrm{H}} \left(\mathrm{V}(\tau(1,\omega_{\mathrm{M}}),\theta^{\mathrm{H}},\omega_{\mathrm{i}}) - \mathrm{V}(\tau(\mathrm{p}^{\mathrm{H}},\omega_{\mathrm{M}}),\theta^{\mathrm{H}},\omega_{\mathrm{i}})\right) \\ & + \mathrm{p}_{\mathrm{L}} \left(\mathrm{V}(\tau(0,\omega_{\mathrm{M}}),\theta^{\mathrm{L}},\omega_{\mathrm{i}}) - \mathrm{V}(\tau(\mathrm{p}^{\mathrm{H}},\omega_{\mathrm{M}}),\theta^{\mathrm{L}},\omega_{\mathrm{i}})\right) \end{aligned} \tag{3.27}$$

Note that there is at least one individual who has positive demand for a think tank, namely the median voter: $\mathrm{D}^{\mathrm{KB}}(\omega_{\mathrm{M}}) > 0$. Under full information she can achieve her highest possible expected utility and is thus trivially better off in the presence of a think tank. Some individuals may not have positive demand for a think tank, however. To see this, note that the first term of equation (3.27) is negative for individuals of type $\omega_{\mathrm{i}} > \overline{\omega}$. Similarly, the second term of equation (3.27) is negative for individuals of type $\omega_{\mathrm{i}} < \underline{\omega}$. There may thus be individuals who are worse off in the presence of a think tank and would prefer that everyone received no further information than their prior. In general, an individual's demand for a think tank is greater if their preferences are more closely aligned with the median voter's preferences.

Let $\Omega_{\mathrm{D^{KB}}} = [\omega_{\mathrm{i}} \mid \mathrm{D}^{\mathrm{KB}}(\omega_{\mathrm{i}}) > 0]$ denote the set of all individuals who receive a net benefit from a think tank. We know that $\Omega_{\mathrm{D^{KB}}}$ is non-empty as it contains at least ω_{M}. Proposition 1 then follows directly.

Proposition 1: *If the expert's bias is known, demand for a think tank with bias $\omega_{\mathrm{TT}} \in [\underline{\omega}, \overline{\omega}]$ exists if and only if $\omega_{\mathrm{E}} \notin [\underline{\omega}, \overline{\omega}]$ and is given by:*

$$\mathrm{D}^{\mathrm{KB}} \;=\; \sum_{\omega_{\mathrm{i}} \in \Omega_{\mathrm{D^{KB}}}} \mathrm{D}^{\mathrm{KB}}(\omega_{\mathrm{i}}) \tag{3.28}$$

Hence we should only expect to observe demand for a think tank if experts are too biased to convey any meaningful information.

3.4.2 Unknown bias

Next, suppose that individuals are uncertain about the expert's preferences. Specifically, I will assume in this section that the expert's bias is $\omega_{\mathrm{E}} < \underline{\omega}$ with probability q, and $\omega_{\mathrm{E}} \in [\underline{\omega}, \omega_{\mathrm{M}}]$ with probability $1 - \mathrm{q}$. Thus, individuals know for certain that the expert has a pro-redistribution bias because $\omega_{\mathrm{E}} < \omega_{\mathrm{M}}$. However, individuals do not know whether the expert's bias is larger or smaller than $\underline{\omega}$ and thus they do not know a priori whether they can believe the expert's message.

With probability q the expert has an "extreme" bias $\omega_{\mathrm{E}} < \underline{\omega}$ and will always send the pro-redistribution message σ^{H}, regardless of the true value of θ. With probability $(1-q)$ the expert has a "moderate" bias $\omega_E \in [\underline{\omega}, \omega_M]$ and will always send a true message about the government's efficiency. This means that if individuals observe the message σ^L, they will understand that the expert must have a "moderate" bias and that they can believe his message. This is because they understand that a pro-redistribution expert who announces that the government's efficiency is low must have a moderate bias and thus must be saying the truth. The logic here is similar to the analysis of Cukierman and Tommasi (1998) that "it takes a Nixon to go to China," that is, unlikely political actions can serve to credibly transmit information. Hence, if individuals observe σ^L, they correctly believe that the government's efficiency is low and update their belief to $b(\sigma^L) = 0$.

On the other hand, if the expert sends the anti-redistribution message σ^H, individuals cannot fully trust the message. They have to account for the possibility that the government's efficiency is actually low, but that the expert has an "extreme" bias and is lying. Thus, individuals use Bayes' rule to update their belief to $b(\sigma^H) = \frac{p^H}{p^H+(1-p^H)q}$.

The equilibrium tax rate will then be given by:

$$\tau(b, \omega_M) \text{ with } b = \begin{cases} 0 \text{ , if } \sigma_E = \sigma^L \\ \frac{p^H}{p^H+(1-p^H)q} \text{ , if } \sigma_E = \sigma^H \end{cases} \tag{3.29}$$

Thus, an individual of type ω_i has the following expected equilibrium utility:

$$\begin{aligned} (1-p^H) \quad & \left[q \, V\left(\tau\left(\frac{p^H}{p^H+(1-p^H)q}, \omega_M\right), \theta^L, \omega_i\right) \right. \\ & \left. \quad + (1-q)V\left(\tau(0, \omega_M), \theta^L, \omega_i\right) \right] \\ + \, p^H \quad & V\left(\tau\left(\frac{p^H}{p^H+(1-p^H)q}, \omega_M\right), \theta^H, \omega_i\right) \end{aligned} \tag{3.30}$$

I will go through each part of equation (30) separately for the sake of clarity. With probability $(1-p^H)$ the government's efficiency is low. If the expert is "extreme," which happens with probability q, he sends a false pro-redistribution message σ^H. Individuals use Bayes' rule to update their beliefs, but they remain uncertain about the true value of θ. Hence, the expert successfully manipulates individual beliefs in this case and distorts the equilibrium tax rate upwards. If the expert is "moderate," he sends a truthful message σ^L and all individuals correctly believe him. As discussed above, this is the only case in which an expert can credibly transmit information and all individuals are perfectly informed.

With probability p^H the government's efficiency is high. In this case the expert will always send the true message σ^L. But because individuals cannot verify whether the expert's message is truthful, they update their belief according to Bayes' rule. In this case the equilibrium tax rate is distorted downwards even though the expert is truthful, because individuals cannot be sure that a pro-redistribution message sent by a pro-redistribution expert is true. Thus, while individuals with "extreme" pro-redistribution biases benefit from distorted beliefs when the government's efficiency is low, they are hurt by the same distortion when the government's efficiency is high and they fail to credibly transmit that information.

Demand for a think tank

Now, suppose that individuals can pool their resources to create a think tank which has a bias $\omega_{TT} \in [\omega_M, \overline{\omega}]$ with probability r, and $\omega_{TT} > \overline{\omega}$ with probability $1 - r$. Such a think tank would represent the interest of individuals with an anti-redistribution bias who are not represented by the expert. Just as the expert's bias is unknown, I assume here that a think tank would similarly have an unknown bias that is "extreme" with some probability r.

In the presence of both an expert and a think tank, individuals form their beliefs on the basis of two messages:

$$b(\sigma_E, \sigma_{TT}) = \Pr[\theta = \theta^H \mid \sigma_E, \sigma_{TT}] \tag{3.31}$$

Since the expert and the think tank have opposing biases, there is a higher chance that individuals will be perfectly informed in equilibrium. Without a think tank, individuals would only know the government's efficiency with certainty if the pro-redistribution expert sends an anti-redistribution message. Now, the opposite may occur as well, when the anti-redistribution think tank sends a pro-redistribution message. By the same logic as above, individuals understand that only a "moderate" think tank would send a pro-redistribution message and thus correctly update their belief to $b = 1$. Thus, the only case in which the median voter is uncertain about θ is when the expert sends $\sigma_E = \sigma^H$ and the think tank sends $\sigma_{TT} = \sigma^L$. In this case, individuals know that one of the senders must be lying, but they do not know which one. Consequently they will update their belief $b(\sigma_E, \sigma_{TT})$ according to Bayes' rule:

$$b(\sigma^L, \cdot) = 0 \tag{3.32}$$

$$b(\cdot, \sigma^H) = 1 \tag{3.33}$$

$$b(\sigma^H, \sigma^L) = \frac{p^H r}{p^H r + (1 - p^H) q} \tag{3.34}$$

After a think tank has been created, expected equilibrium utility for an individual of type ω_i is then given by:

$$\begin{aligned}(1-p^H) \quad & \left[q\, V\left(\tau\left(\frac{p^H r}{p^H r+(1-p^H)q}, \omega_M\right), \theta^L, \omega_i\right)\right. \\ & \left. +(1-q)V\left(\tau(0,\omega_M),\theta^L,\omega_i\right)\right] \\ + p^H \quad & \left[r\, V\left(\tau\left(\frac{p^H r}{p^H r+(1-p^H)q}, \omega_M\right), \theta^H, \omega_i\right)\right. \\ & \left. +(1-r)V\left(\tau(1,\omega_M),\theta^H,\omega_i\right)\right]\end{aligned} \tag{3.35}$$

In the presence of a think tank, the chance that individuals know the true value of θ with certainty increases to $(1-p^H)(1-q)+p^H(1-r)$, namely whenever either the pro-redistribution expert sends an anti-redistribution message or the anti-redistribution think tank sends a pro-redistribution message. In all other cases, individuals remain uncertain about the true value of θ, but the think tank's message lowers their belief b that the government's efficiency is high. Thus, the think tank is effective in persuading the median voter towards a lower tax rate in equilibrium.

Demand for a think tank with unknown bias ω_{TT} for an individual of type ω_i is then given by the difference between equations (35) and (30):

$$\begin{aligned}D^{UB}(\omega_i) \;=\; (1-p^H)q \quad & \left[V\left(\tau\left(\frac{p^H r}{p^H r+(1-p^H)q}, \omega_M\right), \theta^L, \omega_i\right)\right. \\ & \left. - V\left(\tau\left(\frac{p^H}{p^H+(1-p^H)q}, \omega_M\right), \theta^L, \omega_i\right)\right] \\ + p^H r \quad & \left[V\left(\tau\left(\frac{p^H r}{p^H r+(1-p^H)q}, \omega_M\right), \theta^H, \omega_i\right)\right. \\ & \left. - V\left(\tau\left(\frac{p^H}{p^H+(1-p^H)q}, \omega_M\right), \theta^H, \omega_i\right)\right] \\ + p^H(1-r) \quad & \left[V\left(\tau(1,\omega_M),\theta^H,\omega_i\right)\right. \\ & \left. - V\left(\tau\left(\frac{p^H}{p^H+(1-p^H)q}, \omega_M\right), \theta^H, \omega_i\right)\right]\end{aligned} \tag{3.36}$$

In the following I will go through each possible state of the world to shed further light on individual demand for a think tank in this case.

With probability $(1-p^H)q$ the government's efficiency is low but the expert lies and sends a pro-redistribution message σ^H. The think tank sends an opposing message σ^L, thus individuals remain uncertain about the true value of θ. In this case, the think tank's message has a downwards effect on the equilibrium tax rate. The think tank's presence reduces distortions by making individuals' beliefs about the government's efficiency more accurate. All individuals benefit from the think tank's presence except for those who have "extreme" pro-redistribution biases $\omega_i < \underline{\omega}$.

With probability $(1 - p^H)(1 - q)$ the government's efficiency is low and the expert sends a true message σ^L. In this case all individuals know with certainty that $\theta = \theta^L$. The presence of a think tank has no effect on the equilibrium tax rate in this case and therefore this case does not appear in the individual demand equation above.

With probability $p^H r$ the government's efficiency is high but the think tank lies and sends an anti-redistribution message σ^L. The expert sends an opposing message σ^H, thus individuals remain uncertain about the true value of θ. In this case, the think tank's message has a downwards effect on the equilibrium tax rate. However, beliefs become less accurate and are thus distorted by the presence of the think tank. Here, the only individuals who benefit from the think tank's presence are those with "extreme" anti-redistribution biases $\omega_i < \underline{\omega}$, while everyone else is worse off.

With probability $p^H(1 - r)$ the government's efficiency is high and the think tank sends a true message σ^H. In this case all individuals know with certainty that $\theta = \theta^H$ because only a moderate think tank would send a pro-redistribution message. Note that in this case individuals could not have been certain about the government's high efficiency without the presence of a think tank. Thus, the think tank benefits all individuals except for those with "extreme" pro-redistribution biases $\omega_i < \underline{\omega}$.

Let us define $\Omega_{D^{UB}} = [\omega_i \mid D^{UB}(\omega_i) > 0]$ as the set of all individuals who receive a net benefit from a think tank. We know that $\Omega_{D^{UB}}$ is non-empty as it contains at least ω_M. Proposition 2 then follows directly.

Proposition 2: *If the expert's bias is unknown, demand for a think tank with* $r < 1$ *exists if and only if* $q > 0$ *and is given by:* $D^{UB} = \sum_{\omega_i \in \Omega_{D^{UB}}} D^{UB}(\omega_i)$

Thus, if an expert's bias is unknown, demand for a think tank exists if and only if there is a chance $q > 0$ that the expert's bias is so extreme that he cannot convey any meaningful information.

3.5 Conclusion

The analysis in this paper suggests that the existence of a market for think tanks that we observe empirically should raise doubts about the neutrality of academic research. If the electorate fully trusted the information published by academia, we should not expect to observe any demand for think tanks that conduct partisan research.

References

Austen-Smith, David, and John R. Wright. 1992. "Competitive lobbying for a legislator's vote." *Social Choice and Welfare* 9 (3): 229–57. [41]

Bertelli, Anthony M., and Jeffrey B. Wenger. 2009. "Demanding Information: Think Tanks and the US Congress." *British Journal of Political Science* 39 (2): 225–242. [41]

Black, Duncan. 1948. "On the Rationale of Group Decision-making." *Journal of Political Economy* 56 (1): 23–34. [45]

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50 (6): 1431–51. [40, 47]

Cukierman, Alex, and Mariano Tommasi. 1998. "When Does It Take a Nixon to Go to China?" *The American Economic Review* 88 (1): 180–97. [51]

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. "The Economics of Career Concerns, Part I: Comparing Information Structures." *Review of Economic Studies* 66 (1): 183–98. [41]

Downs, Anthony. 1957. "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy* 65 (2): 135–50. [45]

Gans, Joshua S., and Michael Smart. 1996. "Majority voting with single-crossing preferences." *Journal of Public Economics* 59 (2): 219 –237. [45]

McGann, James G. 2016. *The Fifth Estate: Think Tanks, Public Policy, and Governance*. Brookings Institution Press. [39, 40]

Roberts, Kevin W. 1977. "Voting over income tax schedules." *Journal of Public Economics* 8 (3): 329 –340. [40]

Roemer, John E. 1994. "The Strategic Role of Party Ideology When Voters are Uncertain About How the Economy Works." *The American Political Science Review* 88 (2): 327–35. [41]

Saez, Emmanuel, Joel Slemrod, and Seth H. Giertz. 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature* 50 (1): 3–50. [40]