CESifo WORKING PAPERS

9867
2022

Original Version: July 2022
This Version: March 2023

Taxing Moral Agents

Esteban Muñoz Sobrado

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9867*

Taxing Moral Agents

Abstract

Experimental and empirical findings suggest that non-pecuniary motivations play a significant role as determinants of taxpayers' decision to comply with the tax authority and shape their perceptions and assessment of the tax code. By contrast, the canonical optimal income taxation model focuses on material sanctions as the primary motive for compliance. In this paper, I show how taxpayers equipped with semi-Kantian preferences can account for both these non-pecuniary and material motivations. I build a general model of income taxation in the presence of a public good, which agents value morally, and solve for the optimal linear and non-linear taxation problems.

JEL-Codes: H210, H410, D910.

Keywords: optimal income taxation, Kantian agents, prosocial motivations.

Esteban Muñoz Sobrado
Toulouse School of Economics
University of Toulouse Capitole/ France
esteban.munoz@tse-fr.eu

March 28, 2023
I am indebted to Ingela Alger for invaluable guidance and constant support. I acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 789111 - ERC EvolvingEconomics). I am also grateful for comments by Tomer Blumkin, and participants and discussants at the ZEW Public Finance Conference 2022 at the Leibniz Centre for European Economic Research, CESifo Area Conference on Public Economics 2022, the Fairness and the Moral Mind workshop at FAIR Center for Experimental Research on Fairness, the LawPolEcCon 2020 at the Max Plank Institute for Public Goods and Tax Law, ENTER Seminar series at Tilburg University, Inequality, and Rationality, and WIPE 2021 at ECO-SOS, Universitat Rovira i Virgili.

1 Introduction

Tax administration practitioners recognize the importance of non-pecuniary factors as drivers of tax compliance. For instance, Luttmer and Singhal (2014) refer to the following statement by the OECD (2001): "*the promotion of voluntary compliance should be a primary concern of revenue authorities in its principles for good tax administration, and it has highlighted the importance of tax morale more generally* ". This view is consistent with evidence from the World Values Survey (WVS) and European Social Survey (ESS), which indicate that a considerable proportion of citizens perceive tax evasion as being unjustifiable[1] (see Figure 1). Contrastingly, the traditional theoretical analysis of tax evasion (Allingham and Sandmo, 1972) and taxation under asymmetric information (Mirrlees, 1971) focuses on monetary penalties and enforcement as the sole drivers of individual behavior and compliance decisions. While workhorse models of income taxation and income tax evasion view the relationship between the State and its citizens as one of coercion[2], empirical findings show that this cannot be reconciled with high rates of tax compliance observed in some countries (Graetz and Wilde, 1985), nor with experimental findings[3] that find that a considerable proportion of people choose not to evade when playing tax evasion games. More recent findings found in Stantcheva (2021) use large-scale social economics surveys issued to representative U.S. samples and associated experiments to show how social preferences and views of the trustworthiness and scope of government are also crucial drivers of respondents' stance on income tax policy and support for taxes.

In this paper, I consider moral motivations as partial drivers of citizens' sense of civic duty, willingness to pay taxes, and contribute to public goods. In the model, agents consider the role of the government as a provider of public goods when undertaking their compliance

[1] The WVS reports that when asked to rate how justifiable "cheating on taxes if you have a chance" is, 60 percent answer that cheating is never justifiable. In the same vein, 80 percent of the respondents to the ESS "agreed" or "strongly disagreed" with the phrase "citizens should not cheat on their taxes".

[2] According to this coercive view, the taxpayers' main driver to report taxes truthfully is either the possibility of a material sanction (Allingham and Sandmo, 1972) or the design by the Government of an incentive-compatible consumption-leisure bundle (Stiglitz, 1982).

[3] See Alm and Malézieux (2021) for a review of the experimental literature on tax evasion games.

decisions. Particularly, they ask themselves about the hypothetical public good provision that would arise if other members of the society made the same compliance decision as them, holding constant the production function of the government. This is reminiscent of Kant's (1785) categorical imperative; what if a fraction of the population were to act in the same way that I am acting?. It is also compatible with the "social contract" perspective of the State held by Rousseau (1762), which has been previously studied under the label of "reciprocity" between the citizens and the State (Levi, 1989; Besley, 2020).

The model considers agents that have *Homo moralis* preferences. As shown by Alger and Weibull for pair-wise interactions (2013) and then generalized to interactions with infinitely many players (2016), these preferences have strong evolutionary foundations. The model relies on this last generalization and considers an economy with a continuum of agents whose contribution/tax liability funds a global public good, they can be interpreted as agents whose valuation for the public good is constituted by the convex combination of two possible cases: the material public good and a semi-Kantian valuation of the public good. The former valuation is the standard in the literature, it constitutes the "real" public good that a selfish agent derives utility from; the latter considers the material pay-off that she would obtain if all other agents would contribute the same amount that she does, universalizing her actions. *Homo moralis* agents value the public good between these two extremes: they are selfish to some degree, but they also take into account their action in a Kantian sense.

This theoretical setting allows to answer questions regarding the expansion of fiscal capacity in an economy populated with *Homo moralis agents*. More broadly, it also allows to perform normative analysis, considering the problem faced by a utilitarian social planner that maximizes "material" social welfare (absent moral considerations). I consider both the linear and non-linear optimal taxation problems. The results in these two cases write as follows.

First, in the linear income taxation setting, a higher degree of morality is directly linked to an expansion of fiscal capacity: societies with a higher degree of morality can tax income at higher rates and provide more public goods. The public good maximizing income tax that can be implemented by the government increases the degree of morality. *Homo moralis* agents recognize the role played by their taxes at funding a public good and adjust their labor supply accordingly. At a given tax rate, a citizen with higher κ is willing to work more hours if she knows that the income taxes will be used to fund a public good that she values, even if her marginal contribution is atomistic.

Second, in the non-linear income taxation setting, as the government designs the non-linear tax schedule for *Homo moralis* agents an interesting trade-off arises. On the one hand, moral motivations allow the government to collect higher revenues as they relax the incentive constraints of high-ability moral agents. On the other hand, when the government raises the tax paid by low-skilled workers it also crowds out the moral motivation of high-skilled workers, as their Kantian preferences become less stringent at inducing truthful reporting. This result stems from the counter-factual logic employed by Kantian agents: they ask themselves what their utility would be if all the agents of their specific income type were to behave in the same manner as they do. More concretely, when a Kantian agent reports dishonestly to have a lower income and consequently pays a lower income tax, he suffers a utility loss proportional to the difference between the income tax paid by high vs. low-income agents. This means that when low-income agents are paying high taxes, the Kantian concern of high-income types is somewhat "diluted". This also has implications over marginal tax rates of low-income types, which in general increase for low levels of morality and decrease for high morality levels.

At last, for this non-linear taxation environment, I derive a new version of the Samuelson condition which can be directly compared to the one presented by Boadway and Keen (1993). I show that in an economy populated by *Homo moralis* the solution to the problem faced by a utilitarian social planner is such that the sum of marginal rates of substitution between private good and public good consumption is equal to the sum of: (i) the cost of public goods; (ii) the cost of screening, and; (iii) a "moral effect" that affects the provision of public good positively when the net benefit of raising the marginal tax rate for low-skilled agents is high.

Related literature. In the context of public good provision, the possibility of moral considerations has been addressed by authors like Sen (1977), Laffont (1975), and Johansen (1977). For instance the latter states "No society would be viable without some norms and rules of conduct. Such norms and rules are especially necessary for viability in fields where strictly economic incentives are absent and cannot be created. Some degree of honesty in various sorts of communication is one such example, and it might have at least some bearing upon the problem of collective decisionmaking about public goods". More broadly, several forms of intrinsic motivations may be driver citizens' decision to provide public goods[4]. For instance: preferences for honesty (Baiman and Lewis, 1989), social and self-image concerns (Bénabou and Tirole, 2006), or ethical motivations (Laffont, 1975). This paper relates the closest to the latter, which considers the role of Kantian agents in the context of provision of public goods in a large economy, but in the absence of taxation[5].

This work also contributes to the literature on tax morale (Luttmer and Singhal, 2014), which studies several types of non-pecuniary motivations for tax compliance. It provides a new potential motivation for the observed variation in tax morale, and adds a new approach to the list of theories that have been studied by the literature, among those: (i) "warm glow" or impure altruism (Andreoni et al., 1998; Andreoni, 1990; Dwenger et al., 2016); (ii) reciprocity with the state (Levi, 1988; Feld and Frey, 2002; Torgler, 2005; Alm et al., 1993); (iii) peer effects (Besley, 2020);(iv) culture (Kountouris and Remoundou, 2013; DeBacker et al., 2012); and fairness (Bordignon, 1993; Gordon, 1989). In particular, Gordon (1989) proposes an approach that is based on the "Kantian rule" to determine the fair price for the public goods supplied by the state. In this work, individuals consider it fair to pay as much as they would like others to pay. It is assumed that a taxpayer considers it fair to pay the

[4]Empirically, Dwenger et al. (2016) document a high degree of compliance with the German Protestant Church tax that is consistent with a desire to follow the law.

[5]However, other types of ethical rules have been proposed in Economics. For instance, for the case of voting in large elections, Feddersen et al. (2006) and Coate and Conlin (2004) build on the work of Harsanyi (1982; 1992) and study ethical voters as citizens that are "rule utilitarians" that act as a social planner for their group, which results in positive equilibrium turnout rates.

Kantian tax only if they perceive that everyone else is doing the same, and they will revise their desired payment otherwise. My approach differs from this contribution in two ways. Firstly, it is preference-based and does not require the imposition of a "fairness constraint". Secondly, the focus of Gordon (1989) is on the evasion problem, not redistribution[6].

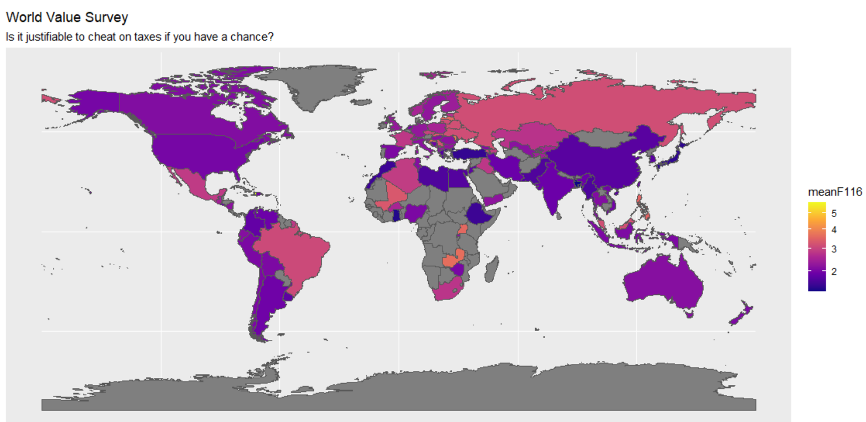

Figure 1: Percentage of people who think cheating on taxes is never justifiable for different countries, WVS. "meanF116" refers refers to the country-average across WVS's waves 1 to 7. A response of 1 asserts that cheating is never justifiable, while higher scores indicate higher justifiability of cheating in taxes.

Finally, this work contributes directly to the literature that considers the role of Kantian ethics in several economic environments. It closely relates to the early contribution of Laffont (1975), who introduces the notion of Kantian behaviour when individuals optimize in an environment with macroeconomic constraints. More particularly, it is the first study of *Homo moralis* preferences in the optimal income taxation setting, and constitutes another application of these preferences in diverse economics environments: Sarkisian (2017, 2021a, 2021b) (team incentives), and Alger and Laslier (2020) and Alger and Laslier (2021) (vot-

[6]Evasion is not explicitly modeled, but rather through incentive constraints, as in Stiglitz (1982)

ing), Eichner and Pethig (2020b) (Piguvian taxation), Eichner and Pethig (2020a) (climate policy), Norman (2020) (the use of fiat money).

The paper is organized as follows: in Section 2 I introduce the baseline economic model. In Section 3 I establish the main results regarding *Homo-moralis* under income homogeneity for both the voluntary contributions benchmark and the linear income taxation environment. Section 4 expands to account for heterogeneity in income and considers the non-linear income taxation case. Section 5 discusses some applications, and Section 6 concludes.

2 The baseline model

The baseline model studies *Homo moralis* agents (citizens) in an economy with a global public good, to which they may contribute (through voluntary contributions or taxes). Agents are atomistic and differ solely in their pre-tax income.

The public good. The economy is populated by an infinite number of agents, each one indexed by i in the (measurable) continuum $I = [0, 1]$. Each agent $i \in I$ contributes a non-negative amount $g_i \geq 0$ to a public good [7]. The public good is produced according to a linear technology:

$$G = \int_I g_i \, di. \tag{1}$$

An important technical observation is that since agents are atomless, the production of the public good is invariant to individual contributions: $\partial G / \partial g_i = 0$ for each $i \in I$.

Preferences. Agents' preferences are *Homo moralis.* This means that they attach some weight to their material utility, which represents their preferences absent any social or moral concerns, while also attaching some weight to a generalized version of Kantian morality. The

[7] While this paper focuses on the case in which g_i corresponds to a tax liability, g_i may generally also correspond to a voluntary contribution.

exact relationship between material utility and moral concerns is clarified in the following paragraphs.

The material utility function. Preferences over material payoffs follow the typical structure studied in the optimal taxation literature [8]: each agent $i \in I$ derives utility from the consumption of the public good G, private consumption x_i, and the number of hours spent working $l_i \in [0,1]$. The *material utility function* is given by the real-valued, differentiable and strictly concave function over the vector (G, x_i, l_i) :

$$U(G, x_i, l_i). \tag{2}$$

I assume that U satisfies the Inada conditions and that agents enjoy the consumption of both the private and the public good ($\partial U/\partial x_i > 0$,w and $\partial U/\partial G > 0$) but dislike working, as it implies spending fewer hours enjoying leisure ($\partial U/\partial l_i < 0$). Henceforth, I use the notation U_m to refer to the partial derivative of U with respect to the m-th entry of the vector (G, x_i, l_i).

The type-structure. Each agent $i \in I$ has a productivity-type $w_n \in \{w_l, w_h\}$, where $w_h \geq w_l$. Productivities can also be interpreted as exogenously determined hourly wages and are distributed across the population according to weights $p_h \in (0,1)$,and $p_l = 1 - p_h$. Whenever, $w_h = w_l$ then model is equivalent to one with only one productivity type. For that special case, I omit the index i and refer to labor supply as $l = l^i(w)$ the labour supply of agent $i \in I$ with productivity $w = w_h = w_l$. Define the budget set of a given agent of type n as:

$$\mathcal{B}(x_n, g_n, l_n) = \{(x_n, g_n, l_n) \in \mathbb{R}^2 \times [0,1] : x_n + g_n \leq l_n \cdot w_n\}, \quad \text{for } n \in \{l, h\}.$$

To convey the main features of *Homo moralis* agents in the baseline model, labor supply will be assumed to be provided inelastically by all agents ($l(w_n) = 1$ for all n). This assumption

[8] E.g: Stiglitz (1982), and Bordignon (1993).

will be then relaxed when addressing the optimal taxation problem.

Welfare criterion, Samuelson is king. Throughout the paper, welfare analysis will be based on the material utility function in equation (2), moreover I assume the planner's material welfare function to be utilitarian. This means that a variant of the Samuelson Rule (Samuelson (1954)) applies as a characterization of the set of Pareto-Optimal allocations. In particular, let labour supply be inelastic at $l_h = l_l = 1$ and denote by $(G^*, x^*(w_n))_{n\in\{l,h\}}$ for the welfare maximizing bundles of public good provision and private consumption.

Proposition 1 (Samuelson Rule)**.** *If the planner is utilitarian and labour supply is inelastic, then the socially optimal level of public good provision and private consumption, denoted* (G^*, x_n^*) *for* $i \in \{l, h\}$*, is such that*

$$\sum_{n\in\{l,h\}} p_n \cdot \frac{U_2(G^*, x_n^*, 1)}{U_1(G^*, x_n^*, 1)} = 1 \tag{3}$$

The proof is in the Appendix. Efficiency in the consumption of public goods requires that the (weighted) sum of marginal rates of substitution between private consumption and consumption of the public good is equal to the marginal rate of transformation between the two goods.

3 Income Homogeneity

In this section, I assume that there is only one income-type $w = w_l = w_h > 0$. In this environment, *Homo-moralis* are defined as followed: a partially Kantian agent takes into account the hypothetical impact that her contribution would have over the global public good if it were to be adopted by some share of the population.

Definition 1. *Homo moralis utilities in a large economy.* Assume that every agent in I has a **degree of morality** $\kappa \in [0, 1]$. Let G denote the global public good. *Homo moralis* preferences over the provision of public good for a given agent $i \in I$ that pays a total tax of $T_i \geq 0$ are given by $U(\mathcal{G}(T_i; G, \kappa), x_i)$, where $\mathcal{G}(T_i; G, \kappa)$ is defined as the *moral valuation over the provision of public good* and is given by:

$$\mathcal{G}(T_i; G, \kappa) = (1 - \kappa) \cdot G + \kappa \cdot T_i. \tag{4}$$

The moral valuation of the public good is a convex combination between G, the real public good which would be the only component valued by a selfish agent T_i, the tax paid by agent i, where the weight attached to the latter is the degree of morality κ.

Note that this definition is silent about the nature of T_i: it can be either a voluntary contribution or a tax liability. In this paper, I examine the latter case and leave the remaining case for an accompanying paper.

3.1 Linear optimal income taxation

In this section, I adapt the baseline model to incorporate a government that funds the public good with the proceeds collected from an income tax. I relax the assumption of inelastic labor supply. Under inelastic labor supply, the government would be always able to achieve first-best outcomes as taxation would not induce any changes in the citizens' utility maximization.

A government selects an income tax $\tau \in [0, 1]$ and uses the proceeds to provide the public good G:

$$G = \tau \int_I y_i \, di, \tag{5}$$

where $y_i = wl_i$ denotes the pre-tax income of agent i at tax rate τ.

In this setting, an agent with *Homo moralis* preferences considers what the public good provision would be, if a share κ of the other agents were to pay the same amount of taxes that they pay. The moral-valuation of the public good of an agent with income y_n is given by:

$$\mathcal{G}(\tau y_i; G, \kappa) = (1-\kappa)G + \kappa \cdot \tau y_i. \tag{6}$$

This expression shows how *Homo moralis* agents perceive a positive utility from paying their taxes to provide a public good. Naturally, this raises the marginal benefit of spending time working: *Homo moralis* agents internalize part of the benefit that their taxable income has on the provision of public goods. For simplicity, below I will write $\mathcal{G}^i$ when referring to $\mathcal{G}(T_i; G, \kappa)$.

The Planner's problem. A utilitarian social planner chooses $\tau \in [0,1]$ and a lump-sum demogrant $b \geq 0$ in order to maximize the sum of material utilities taking the public good production function as given and accounting for the strategic behaviour of its citizens (individual rationality constraint). Mathematically:

$$\max_{(G,\tau)} \int_I U(G, (1-\tau)y_i + b, 1 - \frac{y_i}{w_i})di \tag{7}$$

subject to:

$$G = \tau \int_I y_i(\tau)di, \quad \text{and } \{x_i, l_i\} \in \arg\max U(\mathcal{G}^i, x_i, l_i) \text{ for all } i \in I. \tag{8}$$

Proposition 2. *The solution to the program (7) is such that:*

1. *The agents' maximization problem implies that:*

$$\tau = \frac{1 - \frac{U_3(\cdot)}{w}U_2(\cdot)}{1 - \kappa U_1(\cdot)/U_2(\cdot)} \tag{9}$$

2. *There is a unique optimal tax rate $\tau^*(\kappa) \in [0,1]$.*

3. *At any interior solution, we have that $\frac{\partial \tau^*(\kappa)}{\partial \kappa} > 0$*

Proof. Included in Appendix 6.2.

The optimal tax rate τ weakly increases in the degree of morality κ. This is the consequence of the fact that Kantian moral agents recognize the use of resources that their income tax has as a provider of public goods, and adjust their labor supply to be less sensitive to increases in the optimal income tax. The example below displays how part of the mechanism that yields these results stems from an expansion of fiscal capacity.

Example: expansion of fiscal capacity. Assume that the material utility function of the citizens is separable on leisure of the form $U(G, x_i, l_i) = G^\alpha x_i^{1-\alpha} + \log 1 - l_i$ for all $i \in I$, where $\alpha \in (0, 1)$ measures the preferences for the public good. *Homo moralis* agents decide on leisure-consumption bundles (l_i, x_i) according to:

$$\max_{(l_i, x_i)} \mathcal{G}(\tau w l_i; G, \kappa)^\alpha x_i^{1-\alpha} + \log(1 - l_i) \tag{10}$$
$$\text{subject to:} \quad (l_i, x_i) \in \mathcal{B}(\tau; w),$$

where the budget set above is defined as in (8) and $\mathcal{G}(\tau w l_i; G, \kappa)$ is the moral valuation of the public good in 6 evaluated at $l_i = 1 - y_i/w_i$. In an equilibrium, every agent $i \in I$ maximizes 10 taking τ and G as given. Equilibrium labour supply in this case is given by:

$$\hat{l}_i(\tau, \kappa) = 1 - \frac{(1-\tau)^{1-\alpha}(\gamma\tau)^\alpha}{w((1-\alpha) + \alpha\kappa)}. \tag{11}$$

Equilibrium labour supply follows an inverse U-shaped pattern (Figure 11) with respect to the tax rate τ, meaning that starting from $\tau = 0$, raising taxes increases labour supply for moral agents that value the public good according to (6). However, there exists a threshold value of τ, call it $\tilde{\tau}$, such that $l_i^*(\tilde{\tau}, \kappa) > l_i^*(\tau, \kappa)$ for all $\tau \in [0, 1]$ such that $\tau \neq \tilde{\tau}$.. Moreover, $\tilde{\tau}$ is interior and independent of κ. Equilibrium public good provision is given by:

$$\hat{G}(\kappa; \tau) = \tau \cdot \hat{y}_i(\tau, \kappa) = \tau \cdot w_i \hat{l}_i(\tau, \kappa). \tag{12}$$

Figure 2 shows that the equilibrium public good provision $\hat{G}(\kappa;\tau)$ inherits the inverse U-shaped pattern with respect to the income tax. We can notice a "Laffer-like" pattern in which there exists an interior level of the tax rate τ, be it $\tau^L(\kappa)$ such that $\hat{G}(\kappa;\tau) < \hat{G}(\kappa;\tau^L(\kappa))$ for all $\tau \neq \tau^L(\kappa)$. Moreover, $\tau^L(\kappa)$ is increasing in κ, this suggests that homogeneous societies with higher κ would be able to sustain higher taxes without suffering from a decrease in public good provision.

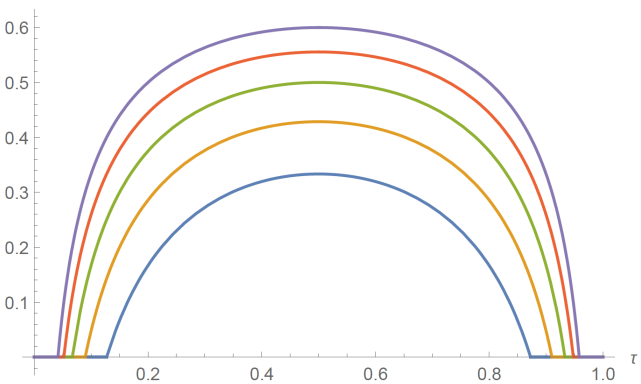

Figure 2: Equilibrium labour supply in an economy of identical agents with $\alpha = 0.5$, $w = 5$ and $\gamma = 1$.

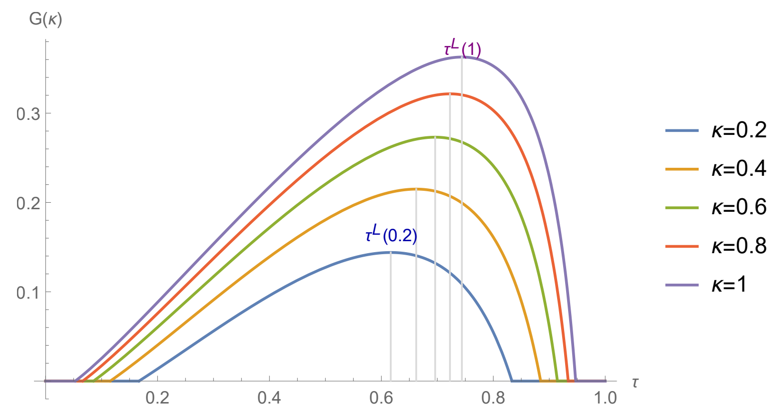

Figure 3: Equilibrium provision of the public good in an economy of identical agents with $\alpha = 0.5$, $w = 5$ and $\gamma = 1$, $\tau^L(\kappa)$ indicates public good maximising "Laffer" rates.

The following section expands these results for the more complex environment in which there is heterogeneity of income types, and agents hold private information on their productivity parameters.

4 Income Heterogeneity

I now solve the non-linear taxation problem a la Mirrlees (1971): each agent private information about his productivity type (how productive they are), while the government knows only the distribution of types and the degree of morality κ, but she cannot observe these characteristics when dealing with a particular agent. The government, however, observes each agent's pre-tax income y. It levies an income tax of $\tau(y)$ and the workers choose consumption and leisure optimally in order to maximize their utility $U\left(G, x, \frac{y}{w_j}\right)$ subject to a private resource constraint $x = y - \tau(y)$ and the government's budget constraint. For convenience, I recur to the following standard notation:

$$U\left(G, x, \frac{y}{w_j}\right) = V^j\left(G, x, y\right), \tag{13}$$

where the index j refers to the agent's true productivity type. Let $\psi(z, w_j)$ denote the marginal rate of substitution between labor and private consumption, where $z = (G, x, y)$:

$$\psi(z, w_j) = \frac{-V_3^j\left(G, x, y\right)}{V_2^j\left(G, x, y\right)}. \tag{14}$$

Assumption 1 (Agent monotonicity or single crossing)**.** *The utility function in* (13) *is such that $\psi^j(z, w_j)$ is a strictly decreasing function of w_j. Or, equivalently, for any z:*

$$\frac{\partial \psi_(z, w_j)}{\partial w_j} < 0. \tag{15}$$

Assumption 1 is the standard *single-crossing condition*. In the same spirit as with equation (14), define the marginal rate of substitution between public good consumption and private good consumption as:

$$\phi(z, w_j) = -\frac{V_1^j\left(G, x, y\right)}{V_2^j\left(G, x, y\right)}. \tag{16}$$

Let there be two types with productivites w_h and w_l: $w_h > w_l$, with proportions p_h, and $p_l = 1 - p_h \in [0, 1]$. The government cannot observe w_j nor l separately. However, it observes that each agent's pre-tax income is given by $y = w_j \cdot l$ and is able to tax it according to the tax function $\tau(y)$. Therefore, each agent's budget set is given by:

$$\mathcal{B}^j = \{(x, y) \in \mathbb{R}_+^2 : x \leq y - \tau(y)\}. \tag{17}$$

The government selects pairs of consumption and pre-tax income (x_n, y_n) for $n \in \{l, h\}$ in order to maximize the utilitarian welfare function subject to the two incentive compatibility and the budget constraint being met. In equilibrium, high-productivity agents choose (x_h, y_h), and low-productivity types choose (x_l, y_l). Hence, the equation for the government's budget constraint is given by:

$$G \leq p_l\tau(y_l) + p_h\tau(y_h) = p_h(y_h - x_h) + p_l(y_l - x_l). \tag{18}$$

As a consequence of their semi-Kantian nature, Homo moralis agents face non-standard incentive constraints which reflect the implications of the Kantian reasoning over their willingness to misreport their true type to the government. More specifically, when a type j chooses the bundle tailored for another type, she internalizes the effect on public good provision that such an action would imply if a share κ of agents of her type were to behave in the same way. Hence, when a type j of corresponding mass p_j selects an income equal to y, she perceives a virtual public good provision equal to:

$$\begin{aligned} \mathcal{G}_i^j &= G + \kappa p_j \left[\tau(y_i) - \tau(y_j)\right] && (19)\\ &= G + \kappa p_j \left[(y_i - x_i) - (y_j - x_j)\right]. && (20) \end{aligned}$$

The above equation is what I refer to as the moral valuation of the public good. A Kantian moral agent values the public good in such a way that he weighs by κ the public good provision that would arise if all agents of his type were to report in the same way under the proposed tax code $\tau(\cdot)$ [9].

This will have an effect on the incentive constraints as they will now write:

$$V^j(\mathcal{G}_j^j, x_j, y_j) \geq V^j(\mathcal{G}_i^j, x_r, y_r), \quad \text{for all } r \neq j. \tag{21}$$

[9]This is a simplification, since the design of $\tau(\dot{)}$ may also serve for redistribution concerns, however, I abstract from this complication in the present exposition.

Noting that $\mathcal{G}_j^j = G$, the government's program hence writes:

$$\begin{aligned}
&\max_{x_h,x_l,y_h,y_l} p_h \cdot V^h(G,x_h,y_h) + p_l \cdot V^l(G,x_l,y_l) \\
&\qquad (BC): \quad p_h \cdot (y_h - x_h) + p_l \cdot (y_l - x_l) \geq G \\
&\qquad (IC_h): \quad V^h(G,x_h,y_h) \geq V^h\left(\mathcal{G}_l^h, x_l, y_l\right) \\
&\qquad (IC_l): \quad V^l(G,x_l,y_l) \geq V^l\left(\mathcal{G}_h^l, x_h, y_h\right). \\
&\qquad (LS): \quad \frac{y_j}{w_r} \leq 1, \quad \text{for all } j,r \in \{l,h\}.
\end{aligned} \tag{22}$$

Proposition 3 (Solution to program (22))**.** . *Assume that the cross derivative between Public Good and leisure is equal to zero, i.e :* $V_{31}(\cdot) = 0$*. Then, the solution to the problem defined in* (22) *for any* $\kappa \in [0,1]$ *is such that:*

1. *If* (IC_h) *binds, then the following conditions hold:*

   (a) *There is* ***no distortion at the top.*** *the marginal tax paid by the high ability type agents still remains equal to zero:*

   $$\psi_h(G,x_h,y_h,w_h) = 1;$$

   (b) *There is* ***distortion at the bottom.*** *Low skilled agents face a lower marginal tax rate, but the marginal tax rate depends on* κ *according to a function* $\alpha(\kappa)$ *such that:*

   $$\psi_l(G,x_l,y_l,w_l) = \alpha(\kappa) < 1, \quad \textit{for } \alpha(\kappa) > 0.$$

2. *If* (IC_l) *binds, then the following conditions hold:*

   (a) ***No distortion at the bottom.*** *the marginal tax faced by the low ability types is equal to zero:*

   $$\psi_l(G,x_l,y_l,w_l) = 1;$$

(b) ***Less intense distortion at the top.*** *High skilled agents face a negative marginal tax rate, which is decreasing in the degree of morality κ according to a function $\gamma(\kappa)$ such that:*

$$\psi_l(G, x_h, y_h, w_h) = \gamma(\kappa) > 1, \quad \textit{for } \gamma(\kappa) < 0.$$

A helpful way to interpret the last proposition is to study the last equation of the proof and consider the expression:

$$\underbrace{\mu \cdot p_l}_{\substack{\text{Marginal benefit of} \\ \text{increasing } \tau(y_l) \text{ in terms} \\ \text{of the public good}}} \quad - \quad \underbrace{\lambda_h \kappa \cdot V_1^h(\mathcal{G}, x_l, y_l)}_{\substack{\text{Marginal cost of} \\ \text{increasing } \tau(y_l) \text{ in terms of} \\ \text{the incentive constraint}}} \quad . \tag{23}$$

The first term constitutes the direct benefit of increasing the tax revenues derived from low-type consumers in terms of the public good. The second term stems from the morality motive embedded in the incentive constraints. This implies that *the planner faces an incentive to distort the marginal tax rate of the less able consumer, but when doing so he also* ***crowds out*** *the moral incentive of the able types.* Recall that moral agents have higher incentives to report truthfully, but such incentives are diluted when misreporting is not very costly in terms of the public good, which is the case when low-ability types face high-income taxes.

When the incentive constraint of the low-ability agents binds the marginal tax rates faced by less able agents are equal to zero, while the marginal tax rate faced by the high-ability individuals is negative: selection constraints require them to work more than they would in a first-best world. Moreover, notice that as the degree of morality κ increases, the marginal tax rate becomes even more negative, this is because to sustain the separating solution, the government must distort the bundle of high-types even further, as moral low-types face a relaxed IC constraint.

4.1 The quasilinear case

The quasilinear case captures the main trade-offs that the planner faces when solving program (22)[10]. Assume that agents can supply L total hours of work[11], and consider the material utility function:

$$U\left(G, x, \frac{y}{w_n}\right) = \theta G + v(x) + \left(L - \frac{y}{w_n}\right), \tag{24}$$

where $v(x)$ is a real-valued twice continuously differentiable function with derivatives $v'(x) > 0$ and $v''(x) < 0$, $\theta \geq 2$, and $h \geq 3$. With this parametrization allows us to characterize several objects presented above. In particular: $\psi_n = \frac{1}{w_n v'(x_n)}$, and $\phi_n(G, x, y, w_n) = \frac{\theta}{v'(x)}$ for $n \in \{l, h\}$. The incentive constraint of the high types writes:

$$v(x_h) - v(x_l) \geq \frac{y_h - y_l}{w_h} - \kappa p_h \theta((y_h - x_h) - (y_l - x_l)).$$

The incentive constraint above is crucial to the result, as the last term at the right-hand-side of the inequality relaxes/tightens the incentive constraint depending on the sign of the term $(y_h - x_h) - (y_l - x_l)$. As we will see, this ambiguity plays an important role in the solution to the planner's problem. Since $\theta \geq 2$, in any solution, the planner decides to set labour supply to its maximum value: $l_n = h$ for all $n \in \{l, h\}$. This consideration, together with the fact that in any solution IC_h yields the no-distortion at the top result result. Let (x_n^{sb}, y_n^{sb}) for $n \in \{l, h\}$ denote the second best solution that solves (22). Then, the following are necessary conditions for (22):

[10] For the interested reader, a solution to the quasilinear case is included in Appendix 6.4

[11] Previously, we used the normalization $L = 1$. Here, we relax this parameter to guarantee interior solutions.

$$v'(x_h^{sb}) = \frac{1}{w_h}, \quad v(x_h^{sb}) - v(x_l^{sb}) = \frac{y_h^{sb} - y_l^{sb}}{w_h} - \kappa p_h \theta((y_h^{sb} - x_h^{sb}) - (y_l^{sb} - x_l^{sb})), \tag{25}$$

$$y_h^{sb} = hw_h, \quad y_l^{sb} = hw_l. \tag{26}$$

These equations implicitly define x_l^{sb}, Figure 4 presents it for some specific parameter values. As can be seen, as for low levels of κ, increases in κ lead to lower levels of x_l compared to the baseline $\kappa = 0$. This effect stems from the fact that the right-hand side of the incentive constraint is now shifted by $-\kappa p_h \theta$ this effect tends to reduce x_l linearly. Now, for low levels of κ, this effect dominates and the principal further distorts x_l downwards to guarantee that high types do not mimic. As we move to the right, we find that there is a $\hat{\kappa}$ such that this effect is reversed. The following proposition fully characterizes it.

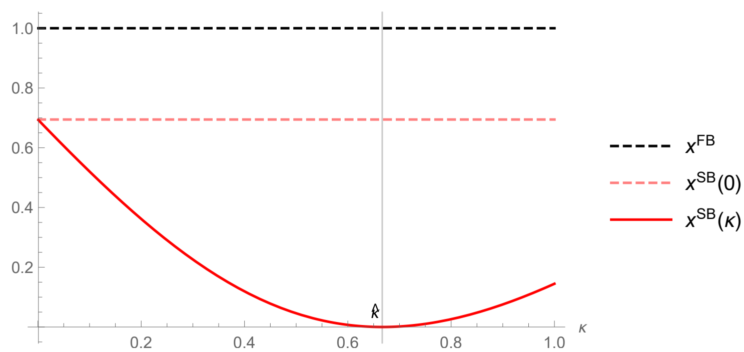

Figure 4: Second best consumption as a function of κ for $v(x) = 2\sqrt{x}$, $\theta = 2$, and $h = 4$.

Proposition 4 (Marginal tax rates in the quasilinear case)**.** *Assume the material utility function is given by 4, then any interior solution to (22), denoted $(x_n^{sb}(\kappa), y_n^{sb}(\kappa))$ for $n \in \{l, h\}$, is such that (25) holds.*

Proof. See Appendix 6.4.

This finding is illustrated by Figure **??** for given parameter values. An entirely selfish agent of low productivity w_l perceives the tax schedule that is implicitly determined by the solution $(x_n^{sb}(\kappa), y_n^{sb}(\kappa))$ to be even further distorted than the baseline case (with $\kappa = 0$)

whenever $\kappa < \hat{\kappa}$, and such effect would, however, be diminished for $\kappa > \hat{\kappa}$. The intuition of this result lies on the behaviour of the incentive constraint and it's effect over the consumption of the low-type that was discussed above. Increasing the degree of morality leads to surprising non-linearities on marginal tax rates once we consider heterogeneous income levels: low levels of morality may induce higher marginal taxes on low types, while this need not be the case for high levels of morality. Next, I characterize the solution to problem (22) for any general utility function. Some of these intuitions still hold, but the derivations are far more involved.

Figure 5 summarizes the result. If κ is low, the principal finds it profitable to raise marginal taxes of low types without incurring a significant incentive costs: I call this the "exploitative effect". On the other hand, if κ is high, it becomes very costly to provide incentives to high-types when marginal taxes are high for low-types (see inequality (59)): I call this, the "moral incentive effect"

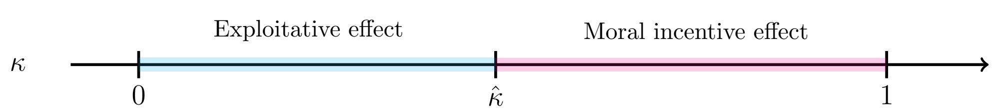

Figure 5: Morality parameter and marginal tax rate of low-ability types.

4.2 On the optimal level of public good provision

Following the approach proposed by Boadway and Keen (1993), I obtain a formula for the distortion in the provision of public goods, and disentangle the part of this effect that stems from the incentive compatibility constraint from the part that is due to the morality motive. For the sake of reducing the length of the notation, I denote the utility of the mimicker as:

$$\hat{V}^h = V^h(\mathcal{G}_l^h, x_n, y_l) \tag{27}$$

Focus on the condition of optimality for the public good given in the proof of Proposition 3. We can add and subtract $\lambda_h \cdot \hat{V}_2^h \left(\frac{V_1^l}{V_1^h}\right)$ and obtain the following:

$$\frac{\partial \mathcal{L}}{\partial G} = \left((1-p_h)V_2^l - \lambda_h \hat{V}_2^h\right) \cdot \frac{V_1^l}{V_2^l} + (p_h + \lambda_h)V_1^h + \lambda_h \hat{V}_2^h \left(\frac{V_1^l}{V_2^l} - \frac{\hat{V}_1^h}{\hat{V}_2^h}\right) = 0 \tag{28}$$

We can now substitute for the terms $(1-p_h)V_2^l - \lambda_h \hat{V}_2^h$ and $(p_h + \lambda_h)$ using the optimality conditions for $\{x_l\}$ and $\{x_h\}$ respectively and obtain the following expression:

$$\frac{1}{\mu}\frac{\partial \mathcal{L}}{\partial G} = \left[(1-p_h)\frac{V_1^l}{V_2^l} + p_h \frac{V_1^h}{V_2^h} - 1\right] + \frac{\lambda_h \hat{V}_2^h}{\mu}\left(\frac{V_1^l}{V_2^l} - \frac{\hat{V}_1^h}{\hat{V}_2^h}\right) + \kappa \frac{\hat{V}_2^h \cdot \lambda_h}{\mu}\left[\frac{V_1^h}{V_2^h}\frac{V_1^h}{\hat{V}_2^h} + \frac{V_1^l}{V_2^l}\frac{\left((1-p_h)V_1^h - \hat{V}_1^h\right)}{\hat{V}_2^h}\right] \tag{29}$$

equation (29) gives us the change in social welfare measured in terms of public sector funds given a raise in the public good G. It contains three elements: (i) the direct effect of increasing the provision of the public good net of the cost (which is 1); (ii) the indirect effect of this increase on the incentive compatibility constraints. These first two effects were studied first by Boadway and Keen (1993). The morality motive, however, provides a new component: (iii) the "moral" or "pro-social" motive. This term implies that the change in social welfare when raising the provision of the public good is proportional to the sum of the marginal rate of substitution of high types between the consumption of the public good and the private good $\frac{V_2^h}{V_1^h}$ and the same marginal rate of substitution for the low types $\frac{V_2^l}{V_1^l}$ adjusted by the net cost of attaining the incentive constraint for the low types $\left((1-p_h)V_2^h - \hat{V}_2^h\right)$.

Proposition 5. *If the social planner is utilitarian, the welfare-maximizing public good pro-*

vision is pinned-down by:

$$\sum_{n\in\{l,h\}} p_n \frac{V_2^n}{V_1^n} = \underbrace{1}_{(y)} + \underbrace{\frac{\lambda_h \hat{V}_1^h}{\mu}\left(\frac{\hat{V}_2^h}{\hat{V}_1^h} - \frac{V_2^l}{V_1^l}\right)}_{(ii)} - \underbrace{\kappa \frac{\hat{V}_1^h \cdot \lambda_h}{\mu}\left[\frac{V_2^h}{V_1^h}\frac{V_2^h}{\hat{V}_1^h} + \frac{V_2^l}{V_1^l}\frac{\left((1-p)V_2^h - \hat{V}_2^h\right)}{\hat{V}_1^h}\right]}_{(iii)} \tag{30}$$

Proposition 5 expands the baseline result obtained by Boadway and Keen (1993): the planner's design problem implies that optimality requires that the sum of marginal rates of substitution is equal to (i) the cost of public goods, plus (ii) a term of distortion that stems from the fact that the planner must choose the optimal level of public good while still providing incentives for the high types to report truthfully. However, the morality motive (iii) provides for a new distortion to the Samuel condition above, which is given by the blue term in equation (30). Again, it is proportional to the net gain of an increase of the taxes for the low type agents.

We can interpret (ii) in the following way: provided $\kappa = 0$, when the low ability types value the public good more than the mimicking $\left(\frac{\hat{V}_1^h}{\hat{V}_2^h} < \frac{V_1^l}{V_2^l}\right)$, then the public good should be over-provided with respect to the social optimum given by the Samuelson Rule. The intuition behind this result is that over-provision can be used by the planner as an instrument for redistribution because of its effect on the incentive constraints. The argument is symmetric for the opposite case in which the low-ability types value the public good less than the mimicker.

Now, focus on (iii), for any positive degree of morality $\kappa > 0$, a positive value of the term in brackets would imply that the planner raises the level of provision of the public good. This would happen when either the (a) baseline utility derived of high types *that don't mimic* V_1^h/V_2^h is high, or (b) the *net benefit of raising the marginal tax rate of the low type* $\left((1-p_h)V_1^h - \hat{V}_1^h\right)$ is high. In the natural case in which this net benefit is negative, this yields an attenuation of the over-provision result implied by (ii), as the crowding out effect described in the previous section implies that redistribution through over-provision of the public good would be more costly compared to the baseline.

5 Discussion and application

The model presented in this paper is designed to be as general as possible and can be applied in a variety of economic environments. Some possible applications are outlined in the appendix, while others are left for future research.

Global Public Goods: Energy Conservation, Climate Action. This model is well-suited for examining global public goods, where individual actions have a minimal impact on overall provision. It is interesting to note the repeated calls for individual action in these contexts, despite the negligible effects of such actions. For example, in one of the earliest contributions to this literature, Laffont (1975) raised this issue in regards to energy conservation: "Why should voluntary conservation efforts work if people are selfish maximizers?" A similar argument can be made today for efforts to reduce high carbon-emitting practices that contribute to the public bad of climate change, such as promoting greener lifestyles, diets, and products, and reducing the use of one-use plastics.

Public or Private Provision: The Case for Charitable Contributions. The model can also be used to examine charitable giving, in which individuals derive utility from contributing to a public good, and the government can complement this through taxes and deductions. This application is discussed in Section **??** (work in progress), based on the work of Diamond (2006).

Civic Virtue. Algan and Cahuc (2009) argues that civic virtue plays a critical role in the design of public unemployment insurance. Future work could explore whether the model presented here yields similar predictions when unemployment insurance is considered a public good.

6 Conclusion

Departing from the useful but unlikely assumption that individuals are *exclusively* motivated by their selfish agendas solves some empirical inconsistencies that are regularly found in the literature in public economics. More specifically, assuming that individuals may be partially motivated by a version of Kantian morality, asking themselves if they are acting according to what they would like to be universal behavior across the population, leads to results that may be closer to the empirical findings regarding voluntary contributions on a public good and willingness to pay taxes.

Homo moralis preferences help explain why voluntary contributions to a public good may be positive even if group size is infinitely large. They provide a channel through which agents may partially internalize the cost that they impose on others when free-riding. This implies a higher public good provision in equilibrium than the one achieved when consumers are entirely selfish. Moreover, public good production may be increasing in the degree of morality of such a population.

The same holds for the case in which individuals do not contribute voluntarily, but instead, there exists a government that is in charge of taxing individuals' labor income to finance the production of the public good. *Homo moralis* preferences predict that in such a setting the average income tax rate will increase to finance a higher provision of public good, while marginal tax rates -however- will still attain the *no distortion at the top* property observed in the typical non-linear taxation problems.

At last, a higher degree of morality is directly linked to an expansion of fiscal capacity: societies with a higher degree of morality can tax income at higher rates and provide more public goods. The public good maximizing income tax that can be implemented by the government increases in the degree of morality.

References

Y. Algan and P. Cahuc. Civic virtue and labor market institutions. *American Economic Journal: Macroeconomics*, 1(1):111–45, 2009.

I. Alger and J.-F. Laslier. Homo moralis goes to the voting booth: coordination and information aggregation. 2020.

I. Alger and J.-F. Laslier. Homo moralis goes to the voting booth: a new theory of voter turnout. 2021.

I. Alger and J. W. Weibull. Homo moralis—preference evolution under incomplete information and assortative matching. *Econometrica*, 81(6):2269–2302, 2013. doi: 10.3982/ECTA10637. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA10637.

I. Alger and J. W. Weibull. Evolution and kantian morality. *Games and Economic Behavior*, 98:56 – 67, 2016. ISSN 0899-8256. doi: https://doi.org/10.1016/j.geb.2016.05.006. URL http://www.sciencedirect.com/science/article/pii/S0899825616300410.

M. G. Allingham and A. Sandmo. Income tax evasion: a theoretical analysis. *Journal of Public Economics*, 1(3):323 – 338, 1972. ISSN 0047-2727. doi: http://dx.doi.org/10.1016/0047-2727(72)90010-2. URL http://www.sciencedirect.com/science/article/pii/0047272772900102.

J. Alm and A. Malézieux. 40 years of tax evasion games: a meta-analysis. *Experimental Economics*, 24(3):699–750, 2021.

J. Alm, B. R. Jackson, and M. McKee. Fiscal exchange, collective decision institutions, and tax compliance. *Journal of Economic Behavior and Organization*, 22(3):285 – 303, 1993. ISSN 0167-2681. doi: https://doi.org/10.1016/0167-2681(93)90003-8. URL http://www.sciencedirect.com/science/article/pii/0167268193900038.

J. Andreoni. Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, 100(401):464–477, 1990. ISSN 00130133, 14680297. URL http://www.jstor.org/stable/2234133.

J. Andreoni, B. Erard, and J. Feinstein. Tax compliance. *Journal of Economic Literature*, 36(2):818–860, 1998. ISSN 00220515. URL http://www.jstor.org/stable/2565123.

S. Baiman and B. L. Lewis. An experiment testing the behavioral equivalence of strategically equivalent employment contracts. *Journal of Accounting Research*, 27(1):1–20, 1989.

T. Besley. State capacity, reciprocity, and the social contract. *Econometrica*, 88(4):1307–1335, 2020.

R. Boadway and M. Keen. Public Goods, Self-Selection and Optimal Income Taxation. *International Economic Review*, 34(3):463–478, 1993. ISSN 0020-6598. doi: 10.2307/2527177. URL http://www.jstor.org.proxy.library.cornell.edu/stable/2527177{%}5Cnhttp://www.jstor.org.proxy.library.cornell.edu/stable/pdfplus/10.2307/2527177.pdf?acceptTC=true.

M. Bordignon. A Fairness Approach to Income-Tax Evasion. *Journal of Public Economics*, 52(3):345–362, 1993. ISSN 00472727. doi: Doi10.1016/0047-2727(93)90039-V.

R. Bénabou and J. Tirole. Belief in a Just World and Redistributive Politics*. *The Quarterly Journal of Economics*, 121(2):699–746, 05 2006. ISSN 0033-5533. doi: 10.1162/qjec.2006.121.2.699. URL https://doi.org/10.1162/qjec.2006.121.2.699.

S. Coate and M. Conlin. A group rule-utilitarian approach to voter turnout: Theory and evidence. *American Economic Review*, 94(5):1476–1504, 2004.

J. M. DeBacker, B. T. Heim, and A. Tran. Importing corruption culture from overseas: Evidence from corporate tax evasion in the united states. Working Paper 17770, National Bureau of Economic Research, January 2012. URL http://www.nber.org/papers/w17770.

P. Diamond. Optimal tax treatment of private contributions for public goods with and without warm glow preferences. *Journal of Public Economics*, 90(4-5):897–919, 2006. URL https://EconPapers.repec.org/RePEc:eee:pubeco:v:90:y:2006:i:4-5:p:897-919.

N. Dwenger, H. Kleven, I. Rasul, and J. Rincke. Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in germany. *American Economic Journal: Economic Policy*, 8(3):203–32, August 2016. doi: 10.1257/pol.20150083. URL http://www.aeaweb.org/articles?id=10.1257/pol.20150083.

T. Eichner and R. Pethig. Climate policy and moral consumers. *Scandinavian Journal of Economics, forthcoming*, 2020a.

T. Eichner and R. Pethig. Kantians defy the economists' mantra of uniform pigovian emissions taxes. 2020b.

T. Feddersen, A. Sandroni, et al. Ethical voters and costly information acquisition. *Quarterly Journal of Political Science*, 1(3):287–311, 2006.

L. P. Feld and B. S. Frey. Trust breeds trust: How taxpayers are treated. *Economics of Governance*, 3(2):87–99, Jul 2002. ISSN 1435-6104. doi: 10.1007/s101010100032. URL https://doi.org/10.1007/s101010100032.

J. P. P. Gordon. Individual morality and reputation costs as deterrents to tax evasion. *European Economic Review*, 33(4):797–805, 1989. URL http://EconPapers.repec.org/RePEc:eee:eecrev:v:33:y:1989:i:4:p:797-805.

M. J. Graetz and L. L. Wilde. The economics of tax compliance: fact and fantasy. *National Tax Journal*, 38(3):355–363, 1985.

J. C. Harsanyi. Rule utilitarianism, rights, obligations and the theory of rational behavior. In *Papers in Game Theory*, pages 235–253. Springer, 1982.

J. C. Harsanyi. Game and decision theoretic models in ethics. *Handbook of game theory with economic applications*, 1:669–707, 1992.

L. Johansen. The theory of public goods: Misplaced emphasis? *Journal of Public Economics*, 7(1):147–152, 1977. ISSN 0047-2727. doi: https://doi.org/10.1016/0047-2727(77)90042-1. URL https://www.sciencedirect.com/science/article/pii/0047272777900421.

I. Kant. Grundlegung der metaphysik der sitten, 1785, akad. *A. IV*, 434, 1785.

Y. Kountouris and K. Remoundou. Is there a cultural component in tax morale? evidence from immigrants in europe. *Journal of Economic Behavior and Organization*, 96(C):104–119, 2013. URL https://EconPapers.repec.org/RePEc:eee:jeborg:v:96:y:2013:i:c:p:104-119.

J.-J. Laffont. Macroeconomic constraints, economic efficiency and ethics: An introduction to kantian economics. *Economica*, 42(168):430–437, 1975.

M. Levi. *Of Rule and Revenue*. University of California Press, 1988. ISBN 9780520067509. URL http://www.jstor.org/stable/10.1525/j.ctt1pngtk.

M. Levi. *Of rule and revenue*. University of California Press, 1989.

E. F. P. Luttmer and M. Singhal. Tax morale. *Journal of Economic Perspectives*, 28(4):149–68, November 2014. doi: 10.1257/jep.28.4.149. URL http://www.aeaweb.org/articles?id=10.1257/jep.28.4.149.

J. A. Mirrlees. An Exploration in the Theory of Optimal Taxation. *Review of Economic Studies*, 38(2):175–208, 1971.

T. W. Norman. The evolution of monetary equilibrium. *Games and Economic Behavior*, 122:233–239, 2020.

J.-J. Rousseau. *Du contract social, ou, Principes du droit politique*, volume 3. Chez Marc Michel Rey, 1762.

P. A. Samuelson. The pure theory of public expenditure, 1954. ISSN 00346535.

R. Sarkisian. Team incentives under moral and altruistic preferences: Which team to choose? *Games*, 8(3):37, 2017.

R. Sarkisian. Optimal incentives schemes under homo moralis preferences. *Games*, 12(1):28, 2021a.

R. Sarkisian. Screening teams of moral and altruistic agents. *Games*, 12(4):77, 2021b.

A. K. Sen. Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, pages 317–344, 1977.

S. Stantcheva. Understanding tax policy: How do people reason? *The Quarterly Journal of Economics*, 136(4):2309–2369, 2021.

J. E. Stiglitz. Self-selection and Pareto efficient taxation. *Journal of Public Economics*, 17 (2):213–240, 1982. ISSN 00472727. doi: 10.1016/0047-2727(82)90020-2.

B. Torgler. Tax morale in latin america. *Public Choice*, 122(1):133–157, 2005. URL https://EconPapers.repec.org/RePEc:kap:pubcho:v:122:y:2005:i:1:p:133-157.

Appendix

6.1 Proofs of proposition 1

Proof. The planner's problem writes:

$$\max_{\{x_l,x_h,G\}} p_h \cdot U(G, x_h, 1) + p_l \cdot U(G, x_l, 1) \tag{31}$$

subject to the public good production constraint:

$$\sum_{n\in\{l,h\}} p_n(w_n - x_n) \geq G. \tag{32}$$

and the feasibility constraints:

$$x_l \in [0, w_l] \quad \text{and} \quad x_h \in [0, w_h]. \tag{33}$$

Since U is increasing in both G and x, equation (32) must bind. Therefore the Lagrangian associated to this problem, with associated multipliers μ_1 and μ_2, writes:

$$\begin{aligned}\mathcal{L}(x_h, x_l, \mu, 1) =& p_h \cdot U\left(\sum_{n\in\{l,h\}} p_n(w_n - x_n), x_h, 1\right) + p_l \cdot U\left(\sum_{n\in\{l,h\}} p_n(w_n - x_n), x_l, 1\right) \\ &+ \mu_1(w_1 - x_1) + \mu_2(w_2 - x_2).\end{aligned} \tag{34}$$

The necessary first-order conditions satisfy:

$$\frac{\partial\mathcal{L}(x_h, x_l, \mu)}{\partial x_h} = p_h\left[-p_h U_1(G, x_h, 1) + U_2(G, x_h, 1)\right] + p_l\left[-p_h U_1(G, x_l, 1)\right] - \mu_1 = 0 \tag{35}$$

$$\frac{\partial\mathcal{L}(x_h, x_l, \mu)}{\partial x_l} = p_h\left[-p_l U_1(G, x_h, 1)\right] + p_l\left[-p_l U_1(G, x_l, 1) + U_2(G, x_h, 1)\right] - \mu_2 = 0 \tag{36}$$

At an interior solution $(x_l, x_h) \in (0, w_l) \times (0, w_H)$ we have that $\mu_1 = \mu_2 = 0$, so we can combine the previous equations to obtain $U_2(G, x_h, 1) = p_l U_1(G, x_l, 1) + p_h U_1(G, x_l, 1) = U_2(G, x_l, 1)$, which we can divide by $U_2(G, x_l)$ and $U_2(G, x_h)$ to obtain:

$$\sum_{n\in\{l,h\}} p_n \cdot \frac{U_1(G, x^*(w_n), 1)}{U_2(G, x^*(w_n), 1)} = 1$$

6.2 Proof of Proposition 2

We can write the objetive function of the agent as $U\left(\mathcal{G}^i, (1-\tau)y_i, 1-\frac{y_i}{w_i}\right)$. Hence, the agent optimaility condition writes:

$$\kappa\tau U_1(\cdot) + (1-\tau)U_2(\cdot) - \frac{1}{w_i}U_3(\cdot) = 0$$

We can divide this equation by $U_2(\cdot)$ and solve for the tax rate:

$$\begin{aligned}
\kappa\tau\frac{U_1(\cdot)}{U_2(\cdot)} + (1-\tau) - \frac{1}{w_i}\frac{U_3(\cdot)}{U_2(\cdot)} &= 0 \\
1-\tau\left(1-\kappa\frac{U_1(\cdot)}{U_2(\cdot)}\right) - \frac{1}{w_i}\frac{U_3(\cdot)}{U_2(\cdot)} &= 0 \\
\tau\left(1-\kappa\frac{U_1(\cdot)}{U_2(\cdot)}\right) &= 1 - \frac{1}{w_i}\frac{U_3(\cdot)}{U_2(\cdot)} \\
\tau &= \frac{1-\frac{1}{w_i}\frac{U_3(\cdot)}{U_2(\cdot)}}{\left(1-\kappa\frac{U_1(\cdot)}{U_2(\cdot)}\right)}
\end{aligned}$$

The Government's objective has an associated lagrangian with mutiplier $\lambda > 0$ give by:

$$\mathcal{L}(G, \tau, \lambda) = U\left(G, y(\tau)(1-\tau), 1-\frac{y}{w}\right) + \lambda(G - \tau y(\tau))$$

The First Order Conditions then write:

$$(\tau): \; -U_2(\cdot)(y(\tau)) - \frac{1}{w}U_3(\cdot) = \lambda(y(\tau))$$

$$(G): \; U_1(\cdot) = -\lambda$$

These two conditions imply that together with the solution of the agents' problem imply that:

$$\tau = \frac{1 - y(\tau)\left(\frac{U_1(\cdot)}{U_2(\cdot)} + 1\right)}{\left(1-\kappa\frac{U_1(\cdot)}{U_2(\cdot)}\right)}$$

6.3 Proof of Proposition 3

When IC_h binds, the Lagrangian associated with problem (22) writes:

$$\mathcal{L}(x_h, y_h, x_l, y_l, G) = p_h \cdot V^h(G, x_h, y_h) + p_l \cdot V^l(G, x_l, y_l) + \lambda_h \left(V^h(G, x_h, y_h) - V^h\left(\mathcal{G}_l^h, x_l, y_l\right)\right) \\ + \mu\left(p_h \cdot (y_h - x_h) + p_l \cdot (y_l - x_l) - G\right) \tag{37}$$

Recalling that $\mathcal{G}_l^h = G + \kappa p_h((y_l - x_l) - (y_h - x_h))$, the necessary first order conditions to this problem write:

$$\frac{\partial \mathcal{L}}{\partial x_h} = p_h \cdot V_2^h(G, x_h, y_h) + \lambda_h V_2^h(G, x_h, y_h) - \lambda_h \kappa p_h V_1^h(\mathcal{G}_l^h, x_l, y_l) - \mu \cdot p_h = 0 \tag{38}$$

$$\frac{\partial \mathcal{L}}{\partial x_l} = p_l \cdot V_2^l(G, x_l, y_l) - \lambda_h\left(-\kappa p_h V_1^h\left(\mathcal{G}_l^h, x_l, y_l\right) + V_2^h\left(\mathcal{G}_l^h(y_l), x_l, y_l\right)\right) - \mu \cdot p_l = 0 \tag{39}$$

$$\frac{\partial \mathcal{L}}{\partial y_h} = p_h \cdot V_3^h(G, x_h, y_h) + \lambda_h\left(V_3^h\left(\mathcal{G}_l^h, x_h, y_h\right) + \kappa p_h V_1^h\left(\mathcal{G}_l^h, x_l, y_l\right)\right) + \mu \cdot p_h = 0 \tag{40}$$

$$\frac{\partial \mathcal{L}}{\partial y_l} = p_l \cdot V_3^l(G, x_l, y_l) - \lambda_h\left(V_3^h(G, x_l, y_l) + \kappa p_h V_1^h(\mathcal{G}_l^h, x_l, y_l)\right) + \mu \cdot p_l = 0 \tag{41}$$

$$\frac{\partial \mathcal{L}}{\partial G} = p_h \cdot V_1^h(G, x_h, y_h) + p_l \cdot V_1^l(G, x_l, y_l) + \lambda_h\left(V_1^h(G, x_h, y_h) - V_1^h\left(\mathcal{G}_l^h, x_l, y_l\right)\right) - \mu = 0 \tag{42}$$

Summing up the first and third equations:

$$p_h \cdot V_2^h(G, x_h, y_h) + p_h \cdot V_3^h(G, x_h, y_h) + \lambda_h\left(V_2^h(G, x_h, y_h) + V_3^h(G, x_h, y_h)\right) = 0. \tag{43}$$

Hence we obtain the no distortion at the top result:

$$\psi_h(G, x_h, y_h) = \frac{-V_3^h(G, x_h, y_h)}{V_2^h(G, x_h, y_h)} = 1. \tag{44}$$

Divide the fourth equation by the second one and obtain:

$$\frac{V_3^l(G, x_l, y_l)}{V_2^l(G, x_l, y_l)} = \frac{-\mu \cdot p_l + \lambda_h\left(V_3^h(\mathcal{G}_l^h, x_l, y_l) + \kappa p_h V_1^h(\mathcal{G}_l^h, x_l, y_l)\right)}{\lambda_h\left(V_2^h(\mathcal{G}_l^h, x_l, y_l) - \kappa V_1^h(\mathcal{G}_l^h, x_l, y_l)\right) + \mu \cdot p_l}. \tag{45}$$

We can now multiply both sides by $(\lambda_h \left(V_2^h(\mathcal{G}_l^h, x_l, y_l) - \kappa V_1^h(\mathcal{G}_l^h, x_l, y_l)\right) + \mu \cdot p_l)/V_2^h(G, x_l, y_l)$:

$$\begin{aligned}
&\frac{V_3^l(G, x_l, y_l)}{V_2^l(G, x_l, y_l)} \left(\frac{\lambda_h \left(V_2^h(\mathcal{G}_l^h, x_l, y_l) - \kappa V_1^h(\mathcal{G}_l^h, x_l, y_l)\right) + \mu \cdot p_l}{V_2^h(\mathcal{G}_l^h, x_l, y_l)} \right) \\
&= \frac{-\mu \cdot p_l + \lambda_h \left(V_3^h(\mathcal{G}_l^h, x_l, y_l) + p_h \kappa V_1^h(\mathcal{G}_l^h, x_l, y_l))\right)}{V_2^h(\mathcal{G}_l^h, x_l, y_l)} \\
&= -\frac{\mu \cdot p_l - \lambda_h p_h \kappa V_1^h(\mathcal{G}_l^h, x_l, y_l)}{V_2^h(\mathcal{G}_l^h, x_l, y_l)} + \frac{\lambda_h V_3^h(\mathcal{G}_l^h, x_l, y_l)}{V_2^h(\mathcal{G}_l^h, x_l, y_l)}.
\end{aligned}$$

Rearranging the last equation we obtain:

$$\frac{\mu \cdot p_l - \lambda_h p_h \kappa V_1^h(\mathcal{G}_l^h, x_l, y_l)}{V_2^h(\mathcal{G}_l^h, x_l, y_l)} \left(1 + \frac{V_3^l(G, x_l, y_l)}{V_2^l(G, x_l, y_l)}\right) = \lambda_h \left(\frac{V_3^h(\mathcal{G}_l^h, x_l, y_l)}{V_2^l(G, x_l, y_l)} - \frac{V_3^l(G, x_l, y_l)}{V_2^l(G, x_l, y_l)}\right) \tag{46}$$

The term in brackets on the left-hand side of the last equation constitutes the marginal tax right for the low ability types. Recall that the single crossing assumption asserts that $\psi_h(G, x_l, y_l) < \psi_l(G, x_l, y_l)$ given that $V_{13} = 0$ by assumption.

2. If $(IC)_L$ binds the Lagrangian associated with problem (22) writes:

$$\begin{aligned}
\mathcal{L}(x_h, y_h, x_l, y_l, G) =& p_h \cdot V^h(G, x_h, y_h) + p_l \cdot V^l(G, x_l, y_l) + \lambda_l \left(V^l(G, x_l, y_l) - V^l\left(\mathcal{G}_h^l, x_h, y_h\right)\right) \\
&+ \mu\left(p_h \cdot (y_h - x_h) + p_l \cdot (y_l - x_l) - G\right)
\end{aligned} \tag{47}$$

Recalling that $\mathcal{G}_h^l = G + \kappa p_l((y_h - x_h) - (y_l - x_l))$, the necessary first order conditions to this problem write. The necessary first order conditions to this problem write:

$$\frac{\partial \mathcal{L}}{\partial x_h} = p_h \cdot V_2^h(G, x_h, y_h) + \lambda_l\left(-V_2^l\left(\mathcal{G}_h^l, x_h, y_h\right) + \kappa \cdot p_l V_1^l\left(\mathcal{G}_h^l, x_h, y_h\right)\right) - \mu \cdot p_h = 0 \tag{48}$$

$$\frac{\partial \mathcal{L}}{\partial y_h} = p_h \cdot V_3^h(G, x_h, y_h) + \lambda_l\left(-V_3^l\left(\mathcal{G}_h^l, x_h, y_h\right) - \kappa \cdot p_l V_1^l\left(\mathcal{G}_h^l, x_h, y_h\right)\right) + \mu \cdot p_h = 0 \tag{49}$$

$$\frac{\partial \mathcal{L}}{\partial x_l} = p_l \cdot V_l^2(G, x_l, y_l) + \lambda_l\left(V_2^l(G, x_l, y_l) - \kappa p_l V_1^l\left(\mathcal{G}_h^l, x_h, y_h\right)\right) - \mu \cdot p_l = 0 \tag{50}$$

$$\frac{\partial \mathcal{L}}{\partial y_l} = p_l \cdot V_l^2(G, x_l, y_l) + \lambda_l\left(V_3^l\left(G, x_l, y_l ht\right) + \kappa p_l V_1^l\left(\mathcal{G}_h^l, x_h, y_h\right)\right) + \mu \cdot p_l = 0 \tag{51}$$

$$\frac{\partial \mathcal{L}}{\partial G} = p_h \cdot V_1^h(G, x_h, y_h) + p_l \cdot V_1^l(G, x_l, y_l) + \lambda_l\left(V_1^l(G, x_l, y_l) - V_1^h\left(\mathcal{G}_h^l, x_h, y_h\right) - \mu = 0\right. \tag{52}$$

in the same manner as in the previous proof, summing up the third and fourth equations:

$$\psi_l(G, x_l, y_l) = \frac{-V_3^l(\mathcal{G}, x_l, y_l)}{V_2^l(G, x_l, y_l)} = 1 \tag{53}$$

On the other hand, we can define again $C(\kappa) = -\kappa V_1^l(\mathcal{G}_h^l, x_h, y_h)$, divide the second equation by the first one and obtain:

$$\frac{V_3^h(G, x_h, y_h)}{V_2^h(G, x_h, y_h)} = \frac{-\mu \cdot p_h + \lambda_l \left(V_3^l(\mathcal{G}_h^l, x_h, y_h) - C(\kappa)\right)}{\lambda_l \left(V_2^l(\mathcal{G}_h^l, x_h, y_h) + C(\kappa)\right) + \mu \cdot p_h} \tag{54}$$

Following the same logic of the previous proof, we can now multiply both sides by:

$$(\lambda_l \left(V_3^l(\mathcal{G}, x_h, y_h) + C(\kappa)\right) + \mu \cdot p_h)/V_2^l(\mathcal{G}, x_h, y_h)$$

and obtain:

$$(1 - \psi_h(G, x_h, y_h)) = \frac{\lambda_l V_2^l(\mathcal{G}_h^l, x_h, y_h)}{\mu \cdot p_h + \lambda_l \cdot C(\kappa)} \left(\psi_h\left(\mathcal{G}_h^l, x_h, y_h\right) - \frac{V_3^l(\mathcal{G}_h^l, x_h, y_h)}{V_2^h(G, x_h, y_h)}\right) < 0 \tag{55}$$

As in the previous proof, the term in brackets is negative as long as there is separability between leisure and the consumption of the public good, which yields the desired result.

6.4 Proof of Section 4: The quasilinear case

Assume that agents have utilities of the form:

$$V^j(\mathcal{G}(y_j), x_j, y_j) = A^j(x_j, y_j) + \theta \cdot \mathcal{G}(\kappa; y_j), \quad \text{for } \theta \geq 1. \tag{56}$$

This means that preferences are quasilinear with respect to the public good. Notice that the single-crossing assumption for the low-ability agents in this case writes:

$$\psi_l(w_l) = \frac{-\partial A^h(x_l, y_l)/\partial y_l}{\partial A^h(x_l, y_l)/\partial x_l} < \frac{-\partial A^l(x_l, y_l)/\partial y_l}{\partial A^l(x_l, y_l)/\partial x_l} = \psi_l(w_h). \tag{57}$$

From the previous equation, notice that quasi linearity implies that single crossing is independent from the consumption of the public good. Using the definition of the moral valuation of the public good presented above:

$$V^j(\mathcal{G}(y_n), x_n, y_n) = A^j(x_n, y_n) + \theta \cdot [(1-\kappa) \cdot G + \kappa \cdot p_j \cdot (y_j - x_j)] \,. \tag{58}$$

Equation (58) allows us to write the incentive constraints of the high-ability agents as:

$$A^h(x_h, y_h) - A^h(x_l, y_l) \geq \kappa \cdot p_h \cdot \theta \left[(y_l - x_l) - (y_h - x_h)\right] . \tag{59}$$

The problem faced by an utilitarian planner that has paternalistic preferences over the provision of public good (i.e, she only considers G instead of $\mathcal{G}(\kappa)$ in her objective function):

$$\begin{aligned}
&\max_{x_h, x_l, y_h, y_l} \theta \cdot G + p_h \cdot V^h\left(x_h, y_h\right) + p_l \cdot V^l\left(x_l, y_l\right) \\
&(BC): \quad p_h \cdot (y_h - x_h) + p_l \cdot (y_l - x_l) \geq G \\
&(IC_h): \quad A^h(x_h, y_h) - A^h(x_l, y_l) \geq -\kappa \cdot p_h \cdot \theta\left((y_h - x_h) - (y_l - x_l)\right) \\
&(IC_l): \quad A^l(x_l, y_l) - A^l(x_h, y_h) \geq -\kappa \cdot p_l \cdot \theta\left((y_l - x_l) - (y_h - x_h)\right)
\end{aligned} \tag{60}$$

Assume that one of the two incentive constraints binds and then substitute this in the objective function of the principal. Notice that the problem is strictly increasing in G, therefore the budget constraint (BC) must bind at any solution. Therefore, substitute the budget constraint in the objective function and write the Lagrangian associated with the problem above as a function of x_n and y_n:

$$\begin{aligned}
\mathcal{L}\left(x_h, y_h, x_l, y_l, \lambda_h\right) =& \theta \cdot \left(\sum_{j \in l,h} p_j(y_j - x_j)\right) + p_h \cdot V^h\left(x_h, y_h\right) + p_l \cdot V^l\left(x_l, y_l\right) \\
&+ \lambda_h \left(A^h(x_h, y_h) - A^h(x_l, y_l) + \kappa \cdot p_h \cdot \theta\left((y_h - x_h) - (y_l - x_l)\right)\right)
\end{aligned} \tag{61}$$

The first-order optimality conditions to this problem write:

$$\frac{\partial \mathcal{L}\left(x_h, y_h, x_l, y_l, \lambda_h\right)}{\partial x_h} = -\theta \cdot p_h + p_h \cdot A^h_{x_h} + \lambda_h \left(A^h_{x_h} - \theta \cdot \kappa \cdot p_h\right) = 0 \tag{62}$$

$$\frac{\partial \mathcal{L}\left(x_h, y_h, x_l, y_l, \lambda_h\right)}{\partial y_h} = \theta \cdot p_h + p_h \cdot A^h_{y_h} + \lambda_h \left(A^h_{y_h} + \theta \cdot \kappa \cdot p_h\right) = 0 \tag{63}$$

$$\frac{\partial \mathcal{L}\left(x_h, y_h, x_l, y_l, \lambda_h\right)}{\partial x_l} = -\theta \cdot p_l + p_l \cdot A^l_{x_l} + \lambda_h \left(-A^h_{x_l} + \theta \cdot \kappa \cdot p_h\right) = 0 \tag{64}$$

$$\frac{\partial \mathcal{L}\left(x_h, y_h, x_l, y_l, \lambda_h\right)}{\partial y_l} = \theta \cdot p_l + p_l \cdot A^l_{y_l} + \lambda_h \left(-A^h_{y_l} - \theta \cdot \kappa \cdot p_h\right) = 0 \tag{65}$$

$$\tag{66}$$

The above system allows us to characterize completely the solution to the planner's problem. First. Notice that we adding the two first order conditions yields:

$$\frac{-\partial A^h(x_h, y_h)/\partial y_h}{\partial A^h(x_h, y_h)/\partial x_h} = 1.$$

It follows from the decentralized solution (see proposition 3) that optimality requires that the planner provides an undistorted bundle to the high-ability types: this is the classic *no distortion at the top* result from the contract theory literature.

Next, we can re-arrange the last two equations provided above in order to obtain:

$$\psi_l(w_l) \triangleq \frac{-A^l_{y_l}}{A^l_{x_l}} = \frac{\theta \cdot p_l - \lambda\left(A^h_{y_l} + \theta \cdot \kappa \cdot p_h\right)}{\theta \cdot p_l + \lambda\left(A^h_{x_l} + \theta \cdot \kappa \cdot p_h\right)} \tag{67}$$

In order to ease the manipulation of the previous equation I define the following constants that will allow to handle the last equation easily:

$$v = \frac{\lambda_h A^h_{x_l}}{\theta p_l}, \quad \text{and} \quad K(\kappa) = \theta \cdot \kappa \cdot p_h. \tag{68}$$

We can now rewrite the previous equation as:

$$\psi_l(w_l) \triangleq \frac{-A^l_{y_l}}{A^l_{x_l}} = \frac{1 - v \cdot K(\kappa) + \psi_l(w_h)}{1 + v - vK(\kappa)} \tag{69}$$

By multiplying the previous equation by $1 + v - vK(\kappa)$ and rearranging the result we obtain:

$$(1 - v \cdot K(\kappa))\,(1 - \psi_l(w_l)) = v \cdot (\psi_l(w_l) - \psi_l(w_h)) \tag{70}$$

Recall that the single crossing assumption implies that the term in the numerator is always positive. On the other hand, the quadratic term $\left(1 - \theta\kappa p_l/A^l_{x_l}\right)\left(1 - \lambda A^h_{x_l}\kappa p_h/p_l\right)$ is increasing in κ if and only if $\kappa > \hat{\kappa}(\theta, \lambda_h, p_h, A^l_{x_l} A^h_{x_l})$ where:

$$\hat{\kappa}(\theta, \lambda_h, p_h, A^l_{x_l}, A^l_{x_l}) = \frac{1}{\theta}\frac{A^l_{x_l}}{p_l} + \frac{1}{\lambda_h}\frac{p_l}{p_h}\frac{1}{A^h_{x_l}}.$$