Sobrado, Esteban Muñoz

Working Paper

Taxing Moral Agents

CESifo Working Paper, No. 9867

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Sobrado, Esteban Muñoz (2022) : Taxing Moral Agents, CESifo Working Paper, No. 9867, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/263797

CESifo WORKING PAPERS

9867
2022

July 2022

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 evolutionary 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.

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

May 12, 2022
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.

TAXING MORAL AGENTS

1 Introduction

Tax administration practitioners recognize the importance of non-pecuniary factors as drivers of tax compliance. For instance, Luttmer and Singhal (2014) present 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. Recent findings by Stantcheva (2021) for the case of income taxation show, using large-scale social economics surveys issued to representative U.S. samples and associated experiments, that 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. I borrow from the literature that studies the long-run evolution of human preferences to propose a new workhorse model of income taxation and public good provision that considers both o pecuniary and non-pecuniary motivations simultaneously as determinants of individual behavior. In the model, agents consider the role of the government as a provider of public goods when undertaking their compliance decisions. Particularly, they ask themselves about the hypothetical public good provision where all the other members of the society made the same compliance decision as them, holding constant the production function of the government. This is 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), they have strong evolutionary foundations. It 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

[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.

the public good is constituted by the convex combination of two possible cases: the material public good and a 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 (i.e according to Kant's (1785) categorical imperative; what if a fraction of the population where to act in the same way that I am acting?).

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. I interpret this as an expansion of the State's fiscal capacity. Finally, the welfare-maximizing income tax is also increasing in the degree of morality, meaning that a social planner would also pick higher income taxes the larger the degree of morality. *Homo moralis* agents recognize the role of 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 one hand, moral motivations allow the government to collect higher revenues as they relax the incentive constraints of high-ability moral agents. However, 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 already 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 agents 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 considered by authors like Sen (1977), Laffont (1975), and Johansen (1977) consider the possibility of ethical and moral motivation as drivers of public good provision. 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". This work relates the closest to that of Laffont (1975), who considers agents that reason in a "Kantian" way, meaning they assume that the other agents act as they do, maximizing their utilities under this "macroeconomic" constraint. 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.

More broadly, several forms of intrinsic motivations may be drivers of tax compliance decisions made by citizens [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, by drawing instead from

[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.

the literature that studies the long-run evolution of human preferences to propose a new workhorse model of income taxation and public good provision that considers both of these pecuniary and non-pecuniary motivations simultaneously.

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 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, 1990a; 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); and (iv) culture (Kountouris and Remoundou, 2013; DeBacker et al., 2012).

The model closely relates to the work of Gordon (1989), who considers ethical norms in the form of a "stigma" cost faced by an agent when evading. Also, Bordignon (1993) models an evasion setting in which taxpayers evade depending on her perception of the fairness of the fiscal treatment, this is modeled through a "fairness constraint" that depends on public good supply, the tax rate, and the perceived tax evasion by other players. This last exploration builds on the "Kantian rule" to determine the fair price to be paid for the public good supplied by the state. In this work, an individual considers it fair to pay as much as he would like other individuals to pay, it is assumed that a taxpayer considers it fair to pay his Kantian tax if and only if he perceives that everybody else does the same and that he revises his desired payment otherwise. My approach differs from his contribution in several aspects. First, it is preference-based, which results in my model not requiring the imposition of a "fairness constraint". Second, the focus of Gordon (1989) is devoted to the evasion problem as opposed to the redistribution[5].

These findings are consistent with views expressed by political philosophers and sociologists who have argued that paying taxes corresponds to civic duty that ought to be respected by citizens. For instance, political philosopher George Klosko [6] refers to a "common good principle" according to which "the government of society X, which provides indispensable (and necessary discretionary) public goods and basic social welfare services may take reasonable measures to promote the common good in additional ways, with citizens required to do their fair shares to support its efforts". This is consonant with the position held by important philosophers of the Enlightenment, including Rousseau (1762), and Locke (1690), who viewed civil and political rights as an exchange between duties from the side of the

[5]In particular, here evasion is not modeled explicitly, but instead through incentive constraints, as in Stiglitz (1982).

[6]See Klosko (2004).

citizens with a benevolent government from the side of the rulers. According to this view, agents may be willing to pay their taxes in exchange for services provided by the state.

Finally, this work contributes directly to the literature that considers the role of Kantian ethics in several economic enviroments. It closely relates to the early contribution of Laffont (1975), who introduces the notion of Kantian behaviour when individuals optimize in an enviroment 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) (voting), Eichner and Pethig (2020b) (piguvian taxation), Eichner and Pethig (2020a) (climate policy), Norman (2020) (the use of fiat money).

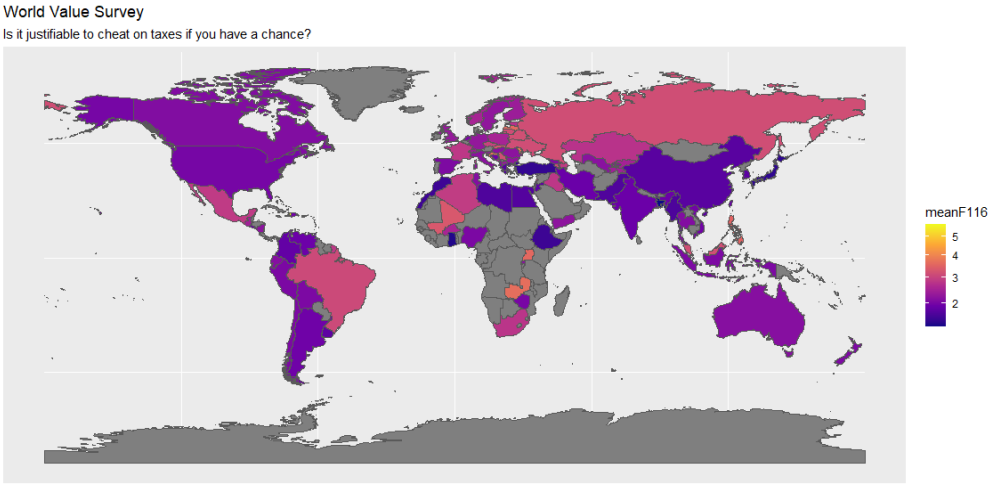

Figure 1: Percentage of people who think cheating on taxes is never justifiable for different countries, WVS. "meanF116" refers refers to the country-average across WVS's waves 1 to 7. A response of 1 asserts that cheating is never justifiable, while higher scores indicate higher justifiability of cheating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

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 atomless 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 $I = [0, 1]$, with associated measure μ. Each agent $i \in I$ may decide to contribute a non-negative amount g_i to a public good denoted by G. The public good is produced according to a linear technology that aggregates over all the individual-level contributions:

$$G = \gamma \int_I g_i \, d\mu(i), \tag{1}$$

where $\gamma \in [-1, 1]$ is a productivity parameter. The typical model of public good provision requires γ to be strictly positive. Instead, here I allow γ to be negative to address the fact that citizens may dislike the Government's choice of public expenditure, indeed transforming it into a public bad. A complementary view is to consider a potentially corrupt Government that can decide to steal a portion of the contributions g_i. Stealing from the fiscal revenues constitutes an act of corruption and produces a public bad, in this case, G accounts for a *net* public good (that is, net of the costs of corruption). 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 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 [7]: 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

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

working l_n. The *material utility function* is given by the real-valued, differentiable and concave function over the vector (G, x_i, l_i) :

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

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 [8] w_n drawn from a discrete set of productivities $W_N = \{w_1, \ldots, w_N\}$, where $N \in \mathbb{Z}^+$ and $w_n > 0$ for all $n = 1, \ldots N$. The proportions of each productivity-type in the population are denoted by p_1, p_2, $\ldots$, p_N, with $\sum_{n=1}^{N} p_n = 1$. Exogenous productivities determine labour supply decisions: i.e $l_n^i = l^i(w_n)$ denotes the labour supply of agent $i \in I$ with productivity w_n for $n = 1, \ldots N$. Henceforth, I abuse notation and omit the sub-index i while keeping the productivity-type sub index $n = 1, \ldots, N$ whenever $N \geq 2$ (this comes without any loss in clarity, since I will focus on type-symmetric equilibria). Define the budget set of a given agent of type n as:

$$\mathcal{B}(x_n, g_n, l(w_n)) = \{(x_n, g_n, l(w_n)) : x_n + g_n \leq l(w_n) \cdot w_n\}, \quad \text{for } n \in 1, \ldots, N.$$

To convey the main features that arise from the model with *Homo moralis* agents, labor supply will be assumed to be provided inelastically by all agents ($l(w_n) = 0$ for all $n \in \{l, h\}$) to obtain some key benchmark results that illustrate the main features of *Homo moralis* in this setting. This assumption 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 simple variant of the Samuelson Rule (Samuelson (1954)) applies as a characterization of the set of Pareto-Optimal allocations. In particular, let $\gamma > 0$ [9], labour supply be inelastic and denote by $(G^*, x^*(w_n))_{n\in\{l,h\}}$ for the welfare maximizing bundles of public good provision and private consumption. Then, a

[8] Productivities can also be interpreted as exogenously determined hourly wages.

[9] When $\gamma \in [-1, 0]$, the unique Pareto-optimal public good provision level is equal to zero, $G^* = 0$.

necessary condition for optimality is given by a

$$\sum_{n\in\{l,h\}} p_n \cdot \frac{U_2(G^*, x^*(w_n))}{U_1(G^*, x^*(w_n)} = \gamma. \tag{3}$$

The detailed derivation of (3) is provided 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.

Equilibrium concept, type-symmetric equilibria. Throughout the paper, I restrict my attention to type-symmetric equilibria in which $(x_n^i, g_n^i, l_n^i) = (x_n, g_n, l_n)$ for all $i \in I$ and all $n \in N$. This means that an equilibrium specifies triplets (x_n, g_n, l_n) for any $n = 1, \dots N$ such that all agents maximize their utilities.

3 Homo moralis under income homonegenity

When $N = 1$, there is only one income-type $w > 0$. In this environment, *Homo-moralis* has a relatively simple definition. In particular, 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 universally adopted. This is put forward in the following definition.

Definition 1 (Homo moralis utilities in a large economy for $N = 1$[10].)**.** *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 contributes $g \geq 0$ are given by $U(\mathcal{G}(g_i; G, \kappa), x_i)$, where $\mathcal{G}(g_i; \mathbf{g}, \kappa)$ is defined as the *moral valuation over the provision of public good* and is given by:

$$\mathcal{G}(g_i; G, \kappa) = \gamma \left[(1 - \kappa) \cdot G + \kappa \cdot g_i\right]. \tag{4}$$

The moral valuation of the public good in the definition above constitutes a convex combination between G, the real public good which would be the only component valued by a selfish agent, and g_i, the contribution of agent i, which is the only behavior that would be considered by a fully Kantian agent, that consider the hypothetical universal adoption of her contribution when deciding over the size of her contribution.

[10]The formal definition and full derivation of *Homo moralis* preferences in a large economy is presented in appendix 7.1.

This definition is silent about the nature of the contribution g_i: g_i can be a voluntary contribution or a tax liability. I explore these two cases extensively throughout the paper.

3.1 Voluntary contributions

Consider the case in which g_i constitutes a voluntary contribution. Suppose that the material utility is quasilinear: $U(G, x_i) = \theta \cdot G + \log x_i$ for all $i \in I$. Denote the aggregate wealth in the economy by $W = \int_I w_i d\mu(i)$, and assume that $\theta\gamma W > 1$[11]. Agents decide on donation-consumption bundles (g_i, x_i) according to:

$$\begin{aligned} &\max_{(g_i,x_i)} \theta \cdot \mathcal{G}(g_i; g_l, g_h, \kappa) + \log x_i \\ &\text{subject to:} \quad (x_i, g_i) \in \mathcal{B}(x_i, g_i; w_i), \end{aligned} \tag{5}$$

where the budget set is given by $\mathcal{B}(x_i, g_i, w) = \{(x_i, g_i) : x_i + g_i \leq w_i\}$. Let $(\hat{x}_i, \hat{g}_i)_{i\in I}$ denote the solution to the program (5) and $\hat{G}$ the resulting equilibrium public good provision.

When $\kappa = 0$, the unique solution to the voluntary contribution problem posed is such that $\hat{g}_i = 0$ for all $i \in I$, which implies that the equilibrium provision of the public good is null, $\hat{G} = 0$. In an economy with atomless agents, no citizen has an incentive to contribute to the public good as incentives to free-ride are infinitely large. If $\kappa > 0$, there are two cases determined by the value of our productivity parameter:

1. if $\gamma \in (0, 1]$, the public good is valued by the citizens, and therefore they derive positive utility from donating marginally to the public good. Hence for all $i \in I$:

$$\hat{g}_i(\kappa) = w - \frac{1}{\theta\gamma\kappa}, \quad \hat{x}_i(\kappa) = \frac{1}{\theta\gamma\kappa}, \quad \text{and } \hat{G}(\kappa) = W - \frac{1}{\theta\gamma\kappa}. \tag{6}$$

2. If $\gamma \in [-1, 0]$, then aggregated contributions result in a public bad. Citizens perceive a net marginal disutility from donating and therefore they decide to abstain from contributing. For all $i \in I$:

$$\hat{g}_i(\kappa) = 0, \quad \hat{x}_i(\kappa) = w, \quad \text{and } \hat{G}(\kappa) = 0. \tag{7}$$

[11] This last condition guarantees that producing the public good is desirable under an utilitarian welfare criterion.

Conditional on the aggregation of contributions resulting in a public good as opposed to a public bad, the equilibrium public good provision with *Homo moralis* agents is increasing in the degree of morality κ and the marginal utility of the public good θ. It is useful to compare the above result with the Pareto-optimal allocation of public good:

$$G^* = \begin{cases} W - \frac{1}{\theta\gamma} & \text{if } \gamma \in (0,1] \\ 0 & \text{if } \gamma \in [-1,0]. \end{cases} \tag{8}$$

The simple comparison of (6) and (8) shows that $\hat{G}(1) = G^*$. When the economy is populated by fully Kantian agents with a degree of morality equal to one, the equilibrium public good provision is exactly equal to the Pareto optimal allocation. The model above is simple but serves as a good benchmark to think about the behavior of *Homo moralis* in a large economy with a public good. The model above can help shed light on two applications.

Remark: warm glow giving. If we restrict our attention to the case in which $\gamma \in (0,1]$, meaning contributions can only be used to provide a net public good, the model above works as a plausible micro foundation for what the literature on "warm glow giving" or "impure altruism" (Andreoni (1990b)). This literature proposed that agents may derive utility from their contributions to a public good, not only because of their potential impact on the amount of public good provided but because of the gift *per se*. This was proposed as an alternative modeling explanation to altruism, which cannot account for empirical regularities in the charitable sector. This is indeed a viable explanation for the setting with atomless agents provided here since citizens with consequentialistic preferences (meaning they optimize over the material utility function in (2)) would never contribute to the public good. Hence, any voluntary contribution observed in such an environment must come from assumptions on the preferences of the citizens.

According to warm glow giving, an agent experiences positive marginal utility from the act of giving. However, the standard model is silent with respect to which motivations may induce agents to experience utility from the act of giving. As the example above shows, *Homo moralis* preferences offer a potential explanation for agents that experience joy from giving. Their decision to contribute is a function of their degree of morality κ, the marginal utility of the public good θ, and the productivity parameter γ.

Application: state capacity and the social contract. An interesting application follows from the general case in which $\gamma \in [-1,1]$. Consider a Government ruled by a potentially corrupt elite. The elite may decide to steal a proportion $\rho \in [0,1]$ of the contribution made by the citizens. Corruption constitutes a public bad, since it erodes democratic institutions, or is

linked to illegal activities. A proportion $(1-\rho)$ of all contributions is devoted to the a gross public good H, where $H = (1-\rho)\int_I g_i d\mu(i)$. The remaining proportion ρ constitutes a public bad B, where $B = \rho\beta\int_I g_i d\mu(i)$, and $\beta \in [0,1]$ is a parameter that measures the damage caused by the appropriation of resources by the elite. If $\beta = 0$, then stealing a fraction ρ of the contributions does not result in any direct harm to the citizens, besides the effect of the reduction in outcome public good H. In general, for any β, stealing a fraction ρ of the contributions causes direct harm of β to the citizens. The net public good, in this case, is given by:

$$G = H - G = \underbrace{[(1-(1+\beta)\rho)]}_{\gamma}\int_I g_i d\mu(i). \tag{9}$$

This equation provides a micro foundation for our parameter γ, in particular, $\gamma = 1-(1+\beta)\rho$ is a decreasing function of the rate of resource-stealing and the harm parameter β. It also means that any elite will not be able to steal more than a fraction $\overline{\rho}(\beta) = 1/(1+\beta)$ of the contributions to the public good according to the equilibrium characterization provided above in equation (6). Crucially, citizen's contributions respond positively to high levels of γ, meaning they provide an upper bound on the elite's kleptocratic drive, This idea, that citizens may respond to their elite's behavior was first proposed by Levi (1988), and has been recently explored closely by Besley (2020). In Appendix 7.4 this application is developed further.

3.2 Linear income taxation

Now, consider the case in which g_i constitutes a tax instead of a voluntary contribution. This distinction is of great importance, since a high proportion of global public goods are not funded voluntarily like in the model proposed above, but are instead provided by governments that raise funds in a coercive manner. The classical example, national defense, fits our state capacity application conveniently: for instance, in times of war citizens may be motivated to pay their fair share of taxes in the proceeds are devoted to defending them against a foreign threat. Other relevant cases include the fight against climate change, or efforts to conserve biodiversity. 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.

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

$$G(\tau) = \gamma \cdot \tau \int_I y(\tau)\, d\mu(i), \tag{10}$$

where $y(\tau)$ denotes the pre-tax income of agent at tax rate τ and $\gamma \in (0, 1]$ [12].

I relax the assumption of inelastic labor supply [13], meaning that now $l(w) \in [0, 1]$ is a decision variable of each agent, pre-tax income then writes: $y(\tau) = w \cdot (1 - l(\tau))$, and the budget set of each agent is given by:

$$\mathcal{B}(\tau; w) = \{(x_i, l_i) \in \mathbb{R} \times [0, 1] : x_i \leq w(1 - \tau)l_i\}. \tag{11}$$

As is typical in the income taxation literature, changes in the income tax affect labor supply decisions that the Government needs to take into account when deciding upon $\tau \in [0, 1]$.

Adapting definition 7.1 to this setting: an agent with *Homo moralis* preferences considers what the outcome public good provision would be, if all the other agents of their type 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:

$$\mathcal{G}(y_n; \kappa, \tau) = \gamma \left[(1 - \kappa)G(\tau) + \kappa \cdot \tau \cdot y_n\right]. \tag{12}$$

The expression above 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.

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

$$\max_{\tau \in [0,1]} \int_I U(G(\tau), x_i(\tau), l_i(\tau)) d\mu(i) \tag{13}$$

[12] The assumption of a non-negative productivity parameter γ is without loss of generality and simplifies the exposition substantially.

[13] 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.

subject to:

$$G(\tau) = \tau \int_I y(\tau) di, \quad \text{and } \{x_n(\tau), l_n(\tau)\} \in \arg\max U(\mathcal{G}, x, l) \text{ for } (x, l) \text{ in } \mathcal{B}(\tau; w). \tag{14}$$

The following proposition characterizes the optimal tax that solves the planner's problem posed in (13). In particular, it shows that the welfare-maximizing tax rate is increasing in the degree of morality, suggesting that societies with higher degrees of morality would lead to higher taxes, provided that the utilitarian solution is a good approximation to the observed income tax rates.

Proposition 1. *Let $\tau^*(\kappa)$ be an interior solution to (13), then:*

$$\frac{\partial \tau^*(\kappa)}{\partial \kappa} \geq 0. \tag{15}$$

Proof. Included in Appendix 7.8. □

The optimal tax rate τ weakly increases in the degree of morality κ. This is the consequence of the fact that 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 $U(G, x_i, l_i) = G^\alpha x_i^{1-\alpha} + \log l_i$ 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:

$$\begin{aligned} &\max_{(l_i, x_i)} \mathcal{G}(l_i; \tau, \kappa)^\alpha x_i^{1-\alpha} + \log(1 - l_i) \\ &\text{subject to:} \quad (l_i, x_i) \in \mathcal{B}(\tau; w), \end{aligned} \tag{16}$$

where the budget set above is defined as in 11 and $\mathcal{G}(l_i; \tau, \kappa)$ is the moral valuation of the public good in 12 evaluated at $y_i = 1 - l_i/w_i$. In an equilibrium, every agent $i \in I$ maximizes 16 taking τ as given. Equilibrium labour supply in this case is given by:

$$\hat{l}_i(\tau,\kappa) = 1 - \frac{(1-t)^{1-\alpha}(\gamma t)^{\alpha}}{w((1-\alpha)+\alpha\kappa)}. \tag{17}$$

Equilibrium labour supply follows an inverse U-shaped pattern (Figure 2) 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 (12). However, there exists a threshold value of τ, call it $\tilde{\tau}$, such that $1 - 1 - l_i^*(\tilde{\tau},\kappa) > 1 - 1 - l_i^*(\tau,\kappa)$ 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:

$$\hat{G}(\kappa;\tau) = \tau \cdot \hat{y}_i(\tau,\kappa) = \tau \cdot w_i \hat{l}_i(\tau,\kappa). \tag{18}$$

A graphical inspection of (18) shows that the equilibrium public good provision $\hat{G}(\kappa;\tau)$ inherits the inverse U-shaped pattern with respect to the income tax (Figure 18). We can notice a "Laffer-like" outcome: there exists an interior level of the tax rate τ, be it $\tau^L(\kappa)$ such that $G(\tau) < G(\tau^L(\kappa))$ 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.

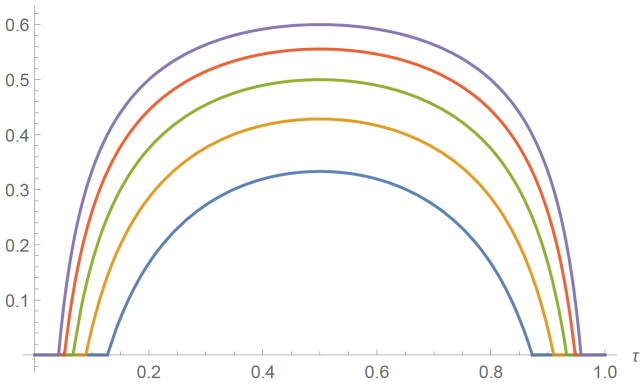

Figure 2: Equilibrium labour supply in an economy of identical agents with $\alpha = 0.5$, $w = 5$ and $\gamma = 1$.

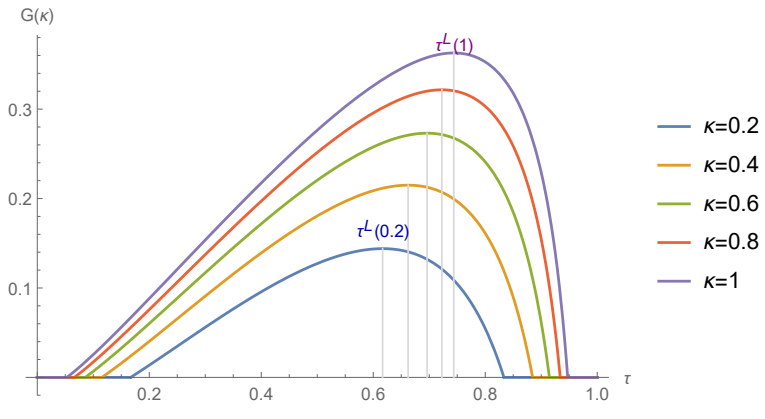

Figure 3: Equilibrium provision of the public good in an economy of identical agents with$\alpha = 0.5$, $w = 5$ and $\gamma = 1$, $\tau^L(\kappa)$ indicates public good maximising "Laffer" rates.

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. This problem is notably more complicated than the one presented above,

not only because the definition of morality is substantially more involved, but also since a utilitarian planner may also be concerned with pure income redistribution.

4 Homo moralis under income heterogeneity

In this section, I set $\gamma = 1$ to focus on the redistribution problem as opposed to the state-building problem explored above. When there is more than a single type, i.e: $N \geq 2$, the definition of *Homo moralis* preferences is slightly more complex, as a theoretical stance is required to be made when defining the reference group that is relevant for the Kantian consideration. In plain words, does a Kantian agent with productivity type w_n consider what the equilibrium public good provision would be if a proportion p_n of the agents in the economy were to contribute the same amount as her (ex post morality), or does she consider the fact that she could have belonged to other groups when estimating the counterfactual Kantian allocation "behind the veil of ignorance" (ex ante morality). Here, I provide the simple definition for the two-type case ($N = 2$) and solve for the quasilinear case. A more general treatment with arbitrary utility functions can be found in Appendix 7.5.

Definition 2 (Homo moralis utilities in a large economy for $N \geq 2$)**.** Assume that every agent in I has a **degree of morality** $\kappa \in [0, 1]$. Let $\bar{g}_h$ and $\bar{g}_l$ denote the average donations for high-types and low-types respectively and $\mathbf{g} = (g_l, g_h)$ the vector of private donations. *Homo moralis* preferences over the provision of public good are given by $U(\mathcal{G}(\mathbf{g}, \kappa), x_n)$, where $\mathcal{G}(\mathbf{g}, \kappa)$ is defined as the *moral valuation over the provision of public good* and is given by:

$$\mathcal{G}^{EA}(g_l, g_h; \kappa, \bar{g}_h, \bar{g}_l) = (1 - \kappa) \cdot (p_h \cdot \bar{g}_h + p_l \cdot \bar{g}_l) + \kappa(p_h \cdot g_h + p_l \cdot g_l) \tag{19}$$

when agents optimize before knowing their types (ex ante optimization). And, denoting by g_n^i the individual contribution made by an n-type, when agents have private information about their types (ex post optimization) the definition writes:

$$\mathcal{G}^{EP}(g_n^i; \kappa, \bar{g}_h, \bar{g}_l) = \begin{cases} p_h \cdot \bar{g}_h + p_l \left[(1 - \kappa) \cdot \bar{g}_l + \kappa \cdot g_l^i\right], & \text{for } n = l. \\ p_l \cdot \bar{g}_l + p_h \left[(1 - \kappa) \cdot \bar{g}_h + \kappa \cdot g_h^i\right], & \text{for } n = h. \end{cases} \tag{20}$$

As equations (19) and (20) show, the two possible definitions for *Homo moralis* preferences share the term $(1 - \kappa)G$, which corresponds to the "real" provision of the public good, while

different on the Kantian component: the ex ante definition in (19) assumes that agents decide on contributions g_h, and g_l before knowing their types, as opposed to the ex post case in (20), in which the reference group is just the realized type $i \in \{l, h\}$.

4.1 Voluntary public good provision

Again, assume that labor supply is inelastic and g_n for $n \in \{l, h\}$ constitute voluntary donations made by low and high-income agents. This subsection established the relationship between the ex ante and ex post definitions provide above, first for the quasilinear case and finally for the general one.

The quasilinear case. Revisiting the quasilinear case, while letting the superscripts EA and EP stand for ex ante and ex post consider the equilibrium with moral agents with degree of morality κ given by $(\hat{x}_n(\kappa), \hat{g}_n(\kappa))$ and resulting public good provision $\hat{G}(\kappa)$ and the Pareto optimal allocations (g_n^*, x_n^*) and provision of the public good G^*. Assume $\gamma \in (0, 1]$:

1. The equilibrium levels and Pareto optimal levels of voluntary contributions are given by:

$$\hat{g}_n^{EP}(\kappa) = w_n - \frac{1}{p_n \theta \gamma \kappa} < \hat{g}_n^{EA}(\kappa) = w_n - \frac{1}{\theta \gamma \kappa} \leq g^*{}_n(\kappa) = w_n - \frac{1}{p_n \theta \gamma}$$

2. The equilibrium and Pareto optimal provision of the public good is given by:

$$\hat{G}^{EP}(\kappa) = W - \frac{2}{\theta \gamma \kappa} < \hat{G}^{EP}(\kappa) = W - \frac{1}{\theta \gamma \kappa} \leq G^* = W - \frac{2}{\theta \gamma}$$

Appendix 7.3 contains the full derivations for the quasilinear case, which conveys the most important conclusion derived from the comparison between ex ante and ex post morality: in particular, the ex post case provides a lower bound for public good provision as equilibrium public good provision is always less than in the ex ante case. This result carries over to more general utility functions.

The general case. The general case is fully developed in Appendix 7.5. Several features of the quasilinear example presented above are also present in the general case for an arbitrary material utility function $U(G, g_n, x_n)$ for $n \in \{l, h\}$. In particular:

1. Both in the ex ante and ex post equilibrium, voluntary contributions are increasing in κ.

2. Public good provision is always weakly higher in the ex ante equilibrium than in the ex post equilibrium: $G^{EA}(\kappa) \geq G^{EP}(\kappa)$.

Figures 4 and 5 evidence points 1 and 2 above for the case of Cobb Douglas utilities. The rest of the paper focuses on the ex post case to draw comparisons with the standard optimal taxation literature. Therefore, point 2 establishes that results included in the sections below should provide a lower bound in terms of public good provision.

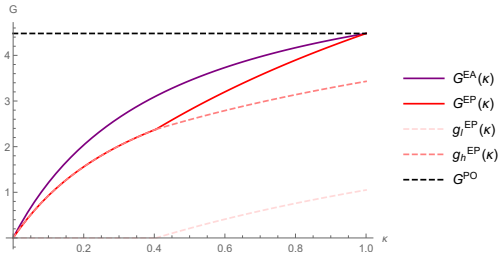

Figure 4: Equilibrium in an economy with moral agents with utilities $\mathcal{G}(\kappa)^{\alpha}x^{1-\alpha}$ for $\alpha = 0.5$, and $\sum_n p_n w_n = 1$.

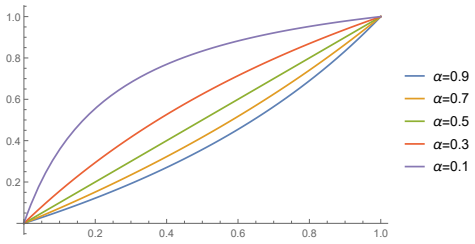

Figure 5: Ratio between Nash equilibrium and Pareto Optimal allocation with moral agents for different values of α.

4.2 Optimal non-linear income taxation

Our last application considers how g_n in definition 20 could represent a non-linear tax liability. For this, set two discrete income types: $w_h > w_l$. In this subsection, I solve the non-linear taxation problem a la Mirrlees (1971). Information is asymmetric; agents know their type (how productive they are), while the designer (i.e, 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, they are each agent's private information.

I study the problem faced by an agent i with productivity $w_{n\in\{l,h\}}$ that pays an income tax of $\tau(y_n)$: she chooses consumption and leisure optimally in order to maximize her utility $U\left(G, x_n, \frac{y_n}{w_n}\right)$ subject to her private resource constraint $x_n = y_n - \tau(y_n)$ and the government's revenue rule $G = \int_n \tau(y_s)ds$. I focus in the case in which there are only two possible productivity levels. In this sense, this model follows the one proposed by Stiglitz (1982). For convenience, I recur to the following standard notation:

$$U\left(G, x_n, \frac{y_n}{w_j}\right) = V^j\left(G, x_n, y_n; w_j\right) = V^j\left(G, x_n, y_n\right) \tag{21}$$

Notice above that the index j refers to the agent's true productivity type. I also draw attention to the fact that the indifference curves of $V(\cdot)$ drawn in the space (x, y) are a function of the ability level w_n and the level of public good consumption G. Let $\psi(G, w_n)$ denote the marginal rate of substitution between pre-tax income and private consumption:

$$\psi_n(G, x_n, y_n; , w_j) = \frac{-V_3^j\left(G, x_n, y_n; w_j\right)}{V_2^j\left(G, x_n, y_n; w_j\right)} = \frac{-U_3\left(G, x_n, \frac{y_n}{w_j}\right)}{w_j \cdot U_2\left(G, x_n, \frac{y_n}{w_j}\right)}. \tag{22}$$

And consider the following assumption.

Assumption 1 (Agent monotonicity or single crossing)**.** *The utility function in* (21) *is such that* $\psi(G, w_n)$ *is a decreasing function of* w_n*. Or, equivalently, for any* G*:*

$$\frac{\partial \psi(G, w)}{\partial w} < 0 \tag{23}$$

or, equivalently, if $\psi(G, w)$ *is not differentiable:*

$$\psi(G, w_h) < \psi(G, w_l) \tag{24}$$

Assumption 1 is commonly referred to as the *single-crossing condition*[14]. In the same spirit as with equation (22), define the marginal rate of substitution between public good consumption and private good consumption as:

$$\phi_j(x_n, G, y_n; , w_j) = -\frac{V_1^j\left(G, x_n, y_n; w_j\right)}{V_2^j\left(G, x_n, y_n; w_j\right)} = \frac{-U_1\left(G, x_n, \frac{y_n}{w_j}\right)}{U_2\left(G, x_n, \frac{y_n}{w_j}\right)} \tag{25}$$

In the absence of taxation each individual's budget constraint is given by: $\mathcal{B}_n' = \{(x_n, l_n) : x_n \leq w_n \cdot l_n\}$. The government cannot observe w_n nor l_n separately. However, it observes that each agent's pre-tax income is given by $y_n = w_n \cdot l_n$ and is able to tax it according to the tax function $\tau(y_n)$. Therefore, each agent's budget set is given by:

$$\mathcal{B}_n = \{(x_n, y_n) : x_n \leq y_n - \tau(y_n)\} \tag{26}$$

Definition 3 (Moral valuation of the public good with an income tax)**.** Let $\tau(y) : \{y_l, y_h\} \to [0, 1]$ be the non-linear tax schedule set by the government. Define the *moral valuation of*

[14]It can be proven (Myles (1995)) that requiring that the private consumption good is not inferior is sufficient for the single-crossing condition to hold.

the public good with an income tax of an agent of type i with degree of morality $\kappa \in [0,1]$ as $\mathcal{G}_n(\kappa)$. Where $\mathcal{G}_n(\kappa)$ is defined as follows:

$$\mathcal{G}_n(\kappa) = (1-\kappa)\cdot G + \kappa \cdot p_n \tau(y_n) \tag{27}$$

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 $\tau(\cdot)$ [15].

The individual rationality of Kantian moral agents. It is useful to fix an arbitrary tax schedule $\tau(y) : \{y_l, y_h\} \to [0,1]$ and study the optimisation problem faced by each agent. This is typically called the "decentralized problem", and it describes the individual behavior that the government should expect after fixing the tax schedule. I characterize it in the following proposition.

Proposition 2 (Decentralization for Homo moralis agents)**.** *If $\kappa \in (0,1]$, and the government commits to a non-linear income tax function given $\tau(y_n)$, then a necessary optimality condition for each agent i of type $n \in \{l,h\}$ is given by:*

$$\tau'(y_n)\left[1 + p_n \cdot \kappa \cdot \phi(\mathcal{G}, w_n)\right] = 1 - \psi(\mathcal{G}, w_n). \tag{28}$$

Moreover, the marginal tax rate at income y_n is given by:

$$\tau'_\kappa(y_n) \triangleq \frac{1-\psi(G,w_n)}{1+p_n\cdot\kappa\cdot\phi(\mathcal{G},w_n)}. \tag{29}$$

Proof. Included in Appendix 7.10. □

Proposition describes how Homo moralis agents equate the marginal rate of substitution between private good consumption and pre-tax income with the after-tax marginal return of working one hour, *adjusted by the moral concern of the virtual externality that this would impose over the provision of the public good.* The agent's optimal response to the tax schedule involves adjusting the marginal tax rate by the factor $(1 + p \cdot \kappa \cdot \phi(\mathcal{G}, w))$; i.e agents take into account the repercussions of their behavior over the overall provision of the public good. The formula above pins down the Homo moralis analog of the standard formula for the marginal tax rate ($\tau'(\cdot) = 1 - \psi(G,w)$), which can be verified by studying equation (29) at $\kappa = 0$:

[15] 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.

$$\tau'_{\kappa=0}(y_n) \triangleq 1 - \psi(G, w_n). \tag{30}$$

The incentive constraints for moral agents. As a consequence of their Kantian concern, 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 agent reports untruthfully, he internalizes the possible effect on public good provision that such a report would imply if all agents of her type were to report in the same way. Hence, when a type j of corresponding mass p_j reports an income equal to r, he perceives a virtual public good provision equal to:

$$\mathcal{G}^j(y_r) = (1 - \kappa) \cdot G + \kappa p_j \cdot \tau(y_r). \tag{31}$$

Moral agents evaluate the material utility that they would obtain if all other agents of their type were to report the same income as they do, this means they are concerned about the implications of their actions in terms of public good provision but neglect the redistributive effects that may be induced by the government's taxation program. This will have an effect over incentive constraints as they will now write [16]:

$$V^j(\mathcal{G}(y_j), x_j, y_j) \geq V^j(\mathcal{G}(y_r), x_r, y_r), \quad \text{for all } r \neq j. \tag{32}$$

The Revelation Principle. Recurring to the revelation principle, I focus on direct mechanism in which agents report truthfully (i.e: incentive-compatible mechanisms). This means, that given the decentralized solution $l_n(w_n, \tau(\cdot))$, one can obtain $y_n(w_n, \tau) = w_n l_n(w_n, \tau(\cdot))$ and $x_n = y_n(w_n, \tau) - \tau(y_n(w_n, \tau))$. Moreover, the solution to the government's problem can be obtained optimising over consumption-income pair (x_n, y_n). Therefore the government's budget constraint can be rewritten as:

$$(BC): \quad G = p_h \cdot (y_h - x_h) + p_l \cdot (y_l - x_l) \tag{33}$$

The Revelation Principle allows us to rewrite the moral valuation of the public good as:

$$\mathcal{G}(\kappa) = (1 - \kappa) \cdot G + \kappa \cdot (p_h \cdot (y_h - x_h) + p_l \cdot (y_l - x_l)). \tag{34}$$

[16] I omit the supra-index j that should correspond to the virtual valuation of the public good $\mathcal{G}^j$, as it coincides with the index j of the function V^j.

Together, equations (33) and (34) allow us to write the planner's problem in the $(x_n, y_n)_{i\in\{l,h\}}$ hyperplane. More concretely, the government selects pairs of consumption and pre-tax income (x_n, y_n) for $i \in \{l, h\}$ in order to maximize the utilitarian welfare function subject the two incentive compatibility and budget constraint being met. The program hence writes:

$$\begin{aligned} \max_{x_h,x_l,y_h,y_l} \; & p_h \cdot V^h\left(G, x_h, y_h\right) + p_l \cdot V^l\left(G, x_l, y_l\right) \\ (BC): \quad & p_h \cdot (y_h - x_h) + p_l \cdot (y_l - x_l) \geq G \\ (IC_h): \quad & V^h\left(\mathcal{G}(y_h), x_h, y_h\right) \geq V^h\left(\mathcal{G}(y_l), x_l, y_l\right) \\ (IC_l): \quad & V^l\left(\mathcal{G}(y_l), x_l, y_l\right) \geq V^l\left(\mathcal{G}(y_h), x_h, y_h\right) \end{aligned} \tag{35}$$

As is standard, I will consider the case in which the incentive constraint of the high types is binding, and leave the other case to the Appendix. I provide a general characterization for any concave utility function U, but first, examine the quasilinear example as it allows to obtain relevant conclusion in terms of the marginal tax rates for the low-productivity agents.

Marginal tax rate for the quasilinear case [17]**.** The quasilinear case captures the main trade-offs that the planner faces when solving program (35). Assume that agents can supply h total hours of work[18], consider the material utility function:

$$U(G, x_n, h - \frac{y_n}{w_n}) = \theta G + v(x_n) + \left(h - \frac{y_n}{w_n}\right), \tag{36}$$

where $v(x_n)$ is a real-valued twice continuously differentiable function with derivatives $v'(x_n) > 0$ and $v''(x_n) < 0$, $\theta \geq 2$, and $h \geq 3$. This parametrization allows us to characterize several objects presented above. In particular: $\psi_n = \frac{1}{w_n v'(x_n)}$, and $\phi_n = \frac{\theta}{v'(x_n)}$ for $n \in \{l, h\}$. This allows to characterize the marginal tax rate as:

$$\tau'_\kappa(y_n) = \frac{1 - \frac{1}{w_n v'(x_n)}}{1 - \kappa p_n \frac{\theta}{v'(x_n)}}. \tag{37}$$

And the incentive constraint of the high types writes:

[17] For the interested reader, a solution to the quasilinear case is included in Appendix 7.10

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

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

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 (35). Then, the following are necessary conditions for (35):

$$v'(x_h^{sb}) = \frac{1}{w_h}, \quad v(x_h^{sb}) - v(x_l^{sb}) = \frac{y_h^{sb} - y_l^{sb}}{w_h} - \kappa p_h \theta((y_h^{sb} - x_h^{sb}) - (y_l^{sb} - x_l^{sb})), \tag{38}$$

$$y_h^{sb} = h w_h, \quad y_l^{sb} = h w_l. \tag{39}$$

These equations implicitly define x_l^{sb}, Figure 6 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.

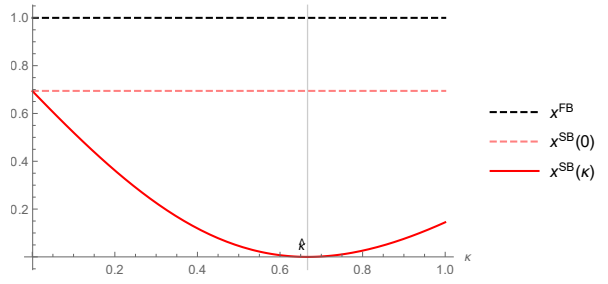

Figure 6: Second best consumption as a function of κ for $v(x) = 2\sqrt{x}$, $\theta = 2$, and $h = 4$.

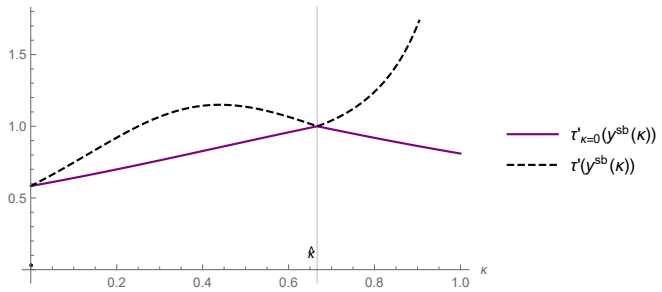

Figure 7: Marginal tax rates for the low-type a function of κ for $v(x) = 2\sqrt{x}$, $\theta = 2$, and $h = 4$.

Proposition 3 (Marginal tax rates in the quasilinear case)**.** *Assume the material utility*

function is given by (36)*, then any interior solution to* (35)*, denoted* $(x_n^{sb}(\kappa), y_n^{sb}(\kappa))$ *for* $n \in \{l, h\}$*, is such that* (38) *holds. Moreover, there exists a threshold* $\hat{\kappa} \in (0, 1)$ *such that: the marginal tax rate that would be experienced by a* $\kappa = 0$ *type (i.e:* $\tau_{\kappa=0}(y^{sb}(\kappa))$*) increases (respectively decreases) in* κ *if* $\kappa \leq \hat{\kappa}$ *(respectively* $\kappa > \hat{\kappa}$*).*

Proof. See Appendix 7.10. □

This finding is illustrated by Figure 7 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 (35) for any general utility function. Some of these intuitions still hold, but the derivations are far more involved.

Proposition 4 (The IC of the high types binds)**.** *When* $\kappa \in (0, 1]$ *and the incentive constraint of the high type is binding, the solution to the problem defined in* (35) *is such that:*

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

$$\psi_h(\mathcal{G}, w_h) = 1;$$

9.2 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:*

$$\psi_l(\mathcal{G}, w_h) = \alpha(\kappa) < 1, \quad \textit{for } \alpha(\kappa) > 0, .;$$

Moreover, the sign of $\alpha'(\kappa)$ *can be positive or negative.*

Proof. The Lagrangian associated with problem (35) writes:

$$\begin{aligned}\mathcal{L}\left(x_h, y_h, x_l, y_l, G\right) =& p_h \cdot V^h\left(G, x_h, y_h\right) + p_l \cdot V^l\left(G, x_l, y_l\right) + \lambda_h\left(V^h\left(\mathcal{G}, x_h, y_h\right) - V^h\left(\mathcal{G}, x_l, y_l\right)\right) \\ &+ \mu\left(p_h \cdot (y_h - x_h) + p_l \cdot (y_l - x_l) - G\right) \end{aligned} \tag{40}$$

The necessary first order conditions to this problem write:

$$\frac{\partial \mathcal{L}}{\partial x_h} = p_h \cdot V_2^h\left(G, x_h, y_h\right) + \lambda_h \left(V_2^h\left(\mathcal{G}, x_h, y_h\right) - \kappa \cdot p_h V_1^h\left(\mathcal{G}, x_h, y_h\right)\right) - \mu \cdot p_h = 0 \tag{41}$$

$$\frac{\partial \mathcal{L}}{\partial x_l} = p_h \cdot V_3^h\left(G, x_h, y_h\right) + \lambda_h \left(V_3^h\left(\mathcal{G}, x_h, y_h\right) + \kappa \cdot p_h V_1^h\left(\mathcal{G}, x_h, y_h\right)\right) + \mu \cdot p_h = 0 \tag{42}$$

$$\frac{\partial \mathcal{L}}{\partial y_h} = p_l \cdot V_2^l\left(G, x_l, y_l\right) + \lambda_h \left(-\kappa p_l V_1^h(\mathcal{G}, x_h, y_h) - V_2^h\left(G, x_l, y_l\right) + \kappa V_1^h(\mathcal{G}, x_l, y_l)\right) - \mu \cdot p_l = 0 \tag{43}$$

$$\frac{\partial \mathcal{L}}{\partial y_l} = p_l \cdot V_3^l\left(G, x_l, y_l\right) + \lambda_h \left(\kappa p_l V_1^h(\mathcal{G}, x_h, y_h) - V_3^h\left(G, x_l, y_l\right) - \kappa V_1^h(\mathcal{G}, x_l, y_l)\right) + \mu \cdot p_l = 0 \tag{44}$$

$$\frac{\partial \mathcal{L}}{\partial G} = p_h \cdot V_1^h\left(G, x_h, y_h\right) + p_l \cdot V_1^l\left(G, x_l, y_l\right) + \lambda_h \left(V_1^h\left(G, x_h, y_h\right) - V_1^h\left(G, x_l, y_l\right)\right) - \mu = 0 \tag{45}$$

summing up the first two equations:

$$p_h \cdot V_2^h\left(G, x_h, y_h\right) + p_h \cdot V_3^h\left(G, x_h, y_h\right) + \lambda_h \left(V_2^h\left(\mathcal{G}, x_h, y_h\right) + V_3^h\left(\mathcal{G}, x_h, y_h\right)\right) = 0 \tag{46}$$

In equilibrium, the virtual valuation of the public good coincides with the real provision of the public good ($G = \mathcal{G}$), hence we obtain the no distortion at the top result:

$$\psi_h(\mathcal{G}, x_h, y_h) = \frac{-V_3^h(\mathcal{G}, x_h, y_h)}{V_2^h(\mathcal{G}, x_h, y_h)} = 1 \tag{47}$$

This means that the high productivity agents' marginal income tax is equal to zero. On the other hand, we can define $C(\kappa) = -\kappa V_1^h(\mathcal{G}, x_l, y_l)$, divide the fourth equation by the the third one and obtain:

$$\frac{V_3^l(\mathcal{G}, x_l, y_l)}{V_2^l(\mathcal{G}, x_l, y_l)} = \frac{-\mu \cdot p_l + \lambda_h \left(V_3^h(\mathcal{G}, x_l, y_l) - C(\kappa)\right)}{\lambda_h \left(V_2^h(\mathcal{G}, x_l, y_l) + C(\kappa)\right) + \mu \cdot p_l} \tag{48}$$

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

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

Rearranging the last equation we obtain:

$$\frac{\mu\cdot p_l+\lambda_h\cdot C(\kappa)}{V_2^h(\mathcal{G},x_l,y_l)}\left(1+\frac{V_3^l(\mathcal{G},x_l,y_l)}{V_2^l(\mathcal{G},x_l,y_l)}\right)=\lambda_h\left(\frac{V_3^h(\mathcal{G},x_l,y_l)}{V_2^l(\mathcal{G},x_l,y_l)}-\frac{V_3^l(\mathcal{G},x_l,y_l)}{V_1^l(\mathcal{G},x_l,y_l)}\right) \tag{49}$$

The term in brackets on the left-hand side of the last equation constitutes the marginal tax right for the low ability types:

$$\tau'(y_l)\left(1+p_h\cdot\kappa\cdot\phi_l(\mathcal{G},w_l)\right)\triangleq\left(1-\psi_l\left(x_l,\mathcal{G},y_l\right)\right)=\frac{\lambda_h V_2^h(\mathcal{G},x_l,y_l)}{\mu\cdot p_l+\lambda_h\cdot C(\kappa)}\left(\psi_l(\mathcal{G},x_l,y_l)-\psi_h(\mathcal{G},x_l,y_l)\right) \tag{50}$$

Recall that the single crossing assumption asserts that $\psi_h(G,x_l,y_l)<\psi_l(G,x_l,y_l)$ this means that term in brackets to the right-hand side of the last equation is positive. Moreover, λ_h and V_1^h are both positive by assumption, so it suffices to study the sign of the denominator on the right:

$$\underbrace{\mu\cdot p_l}_{\substack{\text{Marginal benefit of}\\ \text{increasing } \tau(y_l) \text{ in terms}\\ \text{of the public good}}}-\underbrace{\lambda_h\kappa\cdot V_1^h(\mathcal{G},x_l,y_l)}_{\substack{\text{Marginal cost of}\\ \text{increasing } \tau(y_l) \text{ in terms of}\\ \text{the incentive constraint}}}. \tag{51}$$

□

A helpful way to interpret the last proposition is to look into expression (51). 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.

Appendix 7.11 explores the second case: the solution to the optimal design problem when the incentive constraint of the low-skilled agents is binding. It provides an analogue propo-

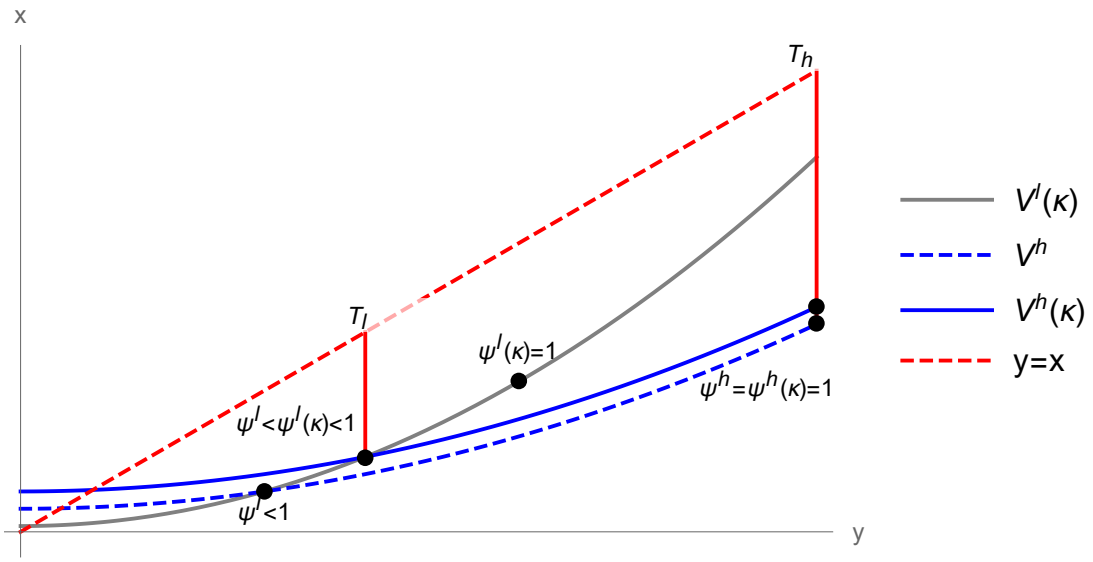

Figure 8: The optimal tax schedule for moral agents is such that high types face a zero marginal tax rate ($\psi_h(\kappa) = 1$), and low-types face a positive marginal tax rate ($\psi_l(\kappa) < 1$). However, the marginal tax rate of low types is lower than the one obtained in an economy populated with selfish agents ($\psi_l < \psi_l(\kappa) < 1$) since the moral concerns of high types imply a shift of the incentive constraint. Low types end up attaining higher bundles of (x_l, y_l).

sition that characterizes this result.

On the optimal level of public good provision

Following the approach prosed by Boadway and Keen (1993), it is possible to obtain a tractable 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:

$$\hat{V}_h = V^h(\mathcal{G}, x_n, y_l) \tag{52}$$

Focus on the condition of optimality for the public good given in the proof of Proposition 12. 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:

$$\frac{\partial \mathcal{L}}{\partial G} = \left((1-p_h)V_2^l - \lambda_h \hat{V}_2^h\right) \cdot \frac{V_1^l}{V_2^l} + (p_h + \lambda_h)V_1^h + \lambda_h \hat{V}_2^h \left(\frac{V_1^l}{V_2^l} - \frac{\hat{V}_1^h}{\hat{V}_2^h}\right) \triangleq 0 \tag{53}$$

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

$$\frac{1}{\mu}\frac{\partial \mathcal{L}}{\partial G} = \left[(1-p_h)\frac{V_1^l}{V_2^l} + p_h \frac{V_1^h}{V_2^h} - 1\right] + \frac{\lambda_h \hat{V}_2^h}{\mu}\left(\frac{V_1^l}{V_2^l} - \frac{\hat{V}_1^h}{\hat{V}_2^h}\right) + \kappa \frac{\hat{V}_2^h \cdot \lambda_h}{\mu}\left[\frac{V_1^h}{V_2^h}\frac{V_1^h}{\hat{V}_2^h} + \frac{V_1^l}{V_2^l}\frac{\left((1-p_h)V_1^h - \hat{V}_1^h\right)}{\hat{V}_2^h}\right] \tag{54}$$

equation (54) 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 $\left((1-p)V_2^h - \hat{V}_2^h\right)$.

Proposition 5. *If the social planner is utilitarian, the welfare-maximizing public good provision is pinned-down by:*

$$\sum_{i\in\{l,h\}} p_n \frac{V_2^i}{V_1^i} = \underbrace{1}_{(y)} + \underbrace{\frac{\lambda_h \hat{V}_1^h}{\mu}\left(\frac{\hat{V}_2^h}{\hat{V}_1^h} - \frac{V_2^l}{V_1^l}\right)}_{(ii)} - \underbrace{\kappa \frac{\hat{V}_1^h \cdot \lambda_h}{\mu}\left[\frac{V_2^h}{V_1^h}\frac{V_2^h}{\hat{V}_1^h} + \frac{V_2^l}{V_1^l}\frac{\left((1-p)V_2^h - \hat{V}_2^h\right)}{\hat{V}_1^h}\right]}_{(iii)} \tag{55}$$

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 (55). 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 applications

With the objective of remaining as general as possible, the model that this paper puts forward is presented in a fairly abstract matter that could be applied in several economic environments. Nevertheless, there are some particular settings and applications in which it could be of relevance. Some of these applications are included in the Appendix and others are left for future research.

Global public goods: energy conservation, climate action. This model is tailored to consider global public goods. In such environments, atomistic individual actions have negligible effects over overall provision. In such context it is puzzling to observe several instances in which a call for individual action is made vigorously. In one of the earliest contributions to this literature, Laffont (1975) puts this idea forward for the case of energy conservation: "For a variety of reasons it is considered in the United States that taxation or rationing to solve the energy problem would be very costly and the government instead asks Americans to voluntarily conserve energy. Why should this work if people are selfish maximizers?". A similar argument can be made today for efforts aimed at reducing individual practices that have high carbon emissions that aggregate into the public bad of irreversible climate change, including the promotion of "greener" lifestyles, diets, and product choices. The same logic applies to the individual reduction of pollutant materials like one-use plastics.

Public or private provision: the case for charitable contributions. The model

can be used as a workhorse model for charitable giving in which agents derive utility from contributing directly to a public good, in an economy in which the government can potentially complement the philanthropic drive from it's citizens via income taxes and charitable deductions. This application is studied in Section 8[19] based on work by Diamond (2006).

Civic virtue. An argument made by Algan and Cahuc (2009) is that civic virtue plays a key role in the design of public insurance against unemployment risk and thus economies with stronger civic virtues are more prone to provide insurance trough unemployment benefits than trough job protection. An exciting avenue for future work is to explore if the model presented here yields similar predictions when unemployment insurance is taken to be a public good.

6 Conclusions

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

[19] This is work in progress.

public goods. The public good maximizing income tax that can be implemented by the government increases in the degree of morality.

7 Appendix

7.1 Homo moralis preferences in a large economy

In this appendix, I follow Alger and Weibull (2016) to formalize *Homo moralis* preferences in an economy with a continuum of agents that derive utility from consumption of public good, a private good, and (possibly) donations to the public good.

Consider an economy populated by a continuum of atomless agents in which each agent derives utility from the consumption of a private good x and the consumption of public good G. The material utility derived by the consumers of this economy is given by:

$$u(x, G). \tag{56}$$

Consumers are exogenously endowed with a monetary amount w_n and can use it for private consumption (x_n) or for donating to the public good (denoting individual donations as d_n). This means that the consumer's budget set is given by:

$$\mathcal{B}_n(w_n) = \{(x_n, g_n) : x_n + g_n \leq w_n\} \tag{57}$$

The production technology for the public good is fairly simple, it consists of the linear aggregation of individual contributions. Hence, if μ denotes the density of types i. The public good is given by:

$$G = \int_I d_n \, d\mu \tag{58}$$

Notice that if the budget constraint binds, we can rewrite the material utility as:

$$\pi(d_n, d_{-i}) = u(w_n - d_n, G(d_n; d_{-i})) \tag{59}$$

which allows us to think about donations d_n as the strategy of individual i against the vector of strategies of all the other agents d_{-i}. We can hence define *Homo moralis preferences* over

the material utility function in (59).

Definition 4 (Homo moralis preferences)**.** Homo moralis preferences with morality type $\mu \in \Delta^{\infty}$ are given by:

$$U(d_n, d_{-i}) = \mathbb{E}\left[\pi(d_n, \mathbf{D})\right], \quad \forall (d_n, d_{-i}) \in (\mathcal{B}_n, \mathcal{B}_{-i}) \tag{60}$$

where $\mathbf{D}$ is a random vector such that with probability μ_m a proportion $m \in [0, 1]$ of the entries of d_{-i} is replaced by d_n, while the other components take their original values.

Now, referring to the terminology of Alger and Weibull (2016) for a given small population share of V-types equal to ϵ, consider the difference between $\mathbb{P}[U; U, \epsilon]$, the conditional probability for an individual of the resident type U that another, uniformly randomly drawn member of his or her group also is of the resident type and $\mathbb{P}[U, V, \epsilon]$, the conditional probability for an individual of the mutant type V that another, uniformly randomly drawn member of his or her group is of the resident type U . This is given by:

$$\phi(\epsilon) = \mathbb{P}[U; U, \epsilon] - \mathbb{P}[U; V, \epsilon] \tag{61}$$

Taking the limit when the mutant type vanished we can define the *index of assortativity* σ as:

$$\sigma = \lim_{\epsilon \to 0} \phi(\epsilon) \tag{62}$$

Assumption 2 (Conditional independence)**.** *For a given mutant who has just been matched, the types of any two other members in her group are statistically independent, at least in the limit as the mutant becomes rare.*

This assumption implies that as $\epsilon \to 0$, the conditional probability of finding other m mutants (q_m^*) follows a binomial distribution (for a finite population with n agents):

$$q_m^* = \frac{(n-1)!}{m!(n-1-m)!} \sigma^m (1-\sigma)^{n-m-1} \tag{63}$$

now, as $n \to \infty$ and for m in the neighbourhood of $n\sigma$, by the de de Moivre–Laplace theorem (normalising population size $n = 1$ and interpreting $m < 1$ as a proportion of

mutants):

$$\mu_m = q_m^* \to \frac{1}{\sqrt{2\pi\sigma(1-\sigma)}} e^{\frac{-(m-\sigma)^2}{2\sigma(1-\sigma)}} = \mathcal{N}\left(\sigma, \sigma(1-\sigma)\right) \tag{64}$$

By the law of large numbers then, the vector $\mathbf{D}$ is such that σ of it's entries are replaced by d_n and $(1-\sigma)$ remain given by d_{-i}. Hence we can rewrite equation (60) as:

$$\begin{aligned} U_\sigma(d_n, d_{-i}) &= \mathbb{E}[\pi(d_n; \sigma d_n + (1-\sigma)d_{-i})] && (65)\\ &= u(di, G(di; \sigma d_n + (1-\sigma)d_{-i})) \quad \text{(def. of } \pi) && (66)\\ &= u(d_n; \sigma d_n + (1-\sigma)d_{-i})) \quad \text{(linear pub. good technology)} && (67) \end{aligned}$$

7.2 Samuelson Condition

Proposition 6 (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* $(G^*, x^*(w_n))$ *for* $i \in \{l, h\}$*, is such that*

$$\sum_{n\in\{l,h\}} p_n \cdot \frac{U_2(G^*, x^*(w_n))}{U_1(G^*, x^*(w_n)} = 1 \tag{68}$$

Proof. The planner's problem writes:

$$\max_{\{x_l, x_h, G\}} p_h \cdot U\left(G, x_h\right) + p_l \cdot U\left(G, x_l\right) \tag{69}$$

subject to the public good production constraint:

$$\sum_{n\in\{l,h\}} p_n(w_n - x_n) \geq G. \tag{70}$$

and the feasibility constraints:

$$x_l \in [0, w_l] \quad \text{and} \quad x_h \in [0, w_h]. \tag{71}$$

Since U is increasing in both G and x, equation (70) must bind. Therefore the Lagrangian

associated to this problem, with associated multipliers μ_1 and μ_2, writes:

$$\begin{aligned}\mathcal{L}\left(x_h, x_l, \mu\right) =& p_h \cdot U\left(\sum_{n\in\{l,h\}} p_n(w_n - x_n), x_h\right) + p_l \cdot U\left(\sum_{n\in\{l,h\}} p_n(w_n - x_n), x_l\right) \\ &+ \mu_1(w_1 - x_1) + \mu_2(w_2 - x_2).\end{aligned} \tag{72}$$

The necessary first-order conditions satisfy:

$$\frac{\partial \mathcal{L}\left(x_h, x_l, \mu\right)}{\partial x_h} = p_h\left[-p_h U_1(G, x_h) + U_2(G, x_h)\right] + p_l\left[-p_h U_1(G, x_l)\right] - \mu_1 = 0 \tag{73}$$

$$\frac{\partial \mathcal{L}\left(x_h, x_l, \mu\right)}{\partial x_l} = p_h\left[-p_l U_1(G, x_h)\right] + p_l\left[-p_l U_1(G, x_l) + U_2(G, x_h)\right] - \mu_2 = 0 \tag{74}$$

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 $U_2(G, x_h) = p_l U_1(G, x_l) + p_h U_1(G, x_l) = U_2(G, x_l)$, which we can divide by $U_2(G, x_l)$ and $U_2(G, x_h)$ to obtain:

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

□

Proposition 6 is a version of the well-known result obtained by Samuelson (1954). It provides a necessary condition for the welfare-maximizing levels of private consumption and public good provision; With this last result, we are well-equipped to compare the private contribution equilibria of the game with *Homo moralis* agents with the first-best normative utilitarian criterion.

7.3 The quasilinear case with $n = 2$

Consider the specific case in which the material utility function is quasilinear: $U(G, x_n) = \theta \cdot G + \log x_n$ for $n \in l, h$, with $\theta(p_h w_h + p_l w_l) > 1$. The last condition guarantees that producing the public good is desirable under an utilitarian welfare criterion. Let's study the voluntary contribution equilibrium for the ex post and ex ante case, in that order. First, an agent of type $n \in \{l, h\}$ decides on contribution-consumption bundles (g_n, x_n) according to the following program:

$$\max_{(g_n,x_n)} \theta \cdot \mathcal{G}^{EP}(g_n;\kappa,\bar{g}_h,\bar{g}_l) + \log x_n \tag{75}$$
$$\text{subject to:} \quad (x_n,g_n) \in \mathcal{B}(x_n,g_n;w_n), \quad \forall n \in \{l,h\}.$$

where the Budget set is given by $\mathcal{B}(x_n,g_n,w_n) = \{(x_n,g_n) : x_n+g_n \leq w_n\}$. Let $(\hat{g}_n^{EP},\hat{g}_n^{EA})_{n\in l,h}$ denote an interior solution program(75) for all agents. It is easy to see that for all $n \in \{l,h\}$:

1. The equilibrium levels of voluntary contributions and consumption are given by:

$$\hat{g}_n^{EP}(\kappa) = w_n - \frac{1}{p_n\theta\kappa}, \text{ and } \hat{x}_n^{EP}(\kappa) = \frac{1}{p_n\theta\kappa}.$$

2. The equilibrium provision of the public good is given by:

$$\hat{G}^{EP}(\kappa) = W - \frac{2}{\theta\kappa},$$

where $W = \sum_{n\in\{l,h\}p_n w_n}$ is the total exogenous wealth of the economy.

Individual contributions are increasing in the degree of morality κ, weighted by the population shares p_n of each productivity type. This yields an outcome equilibrium public good provision $\hat{G}^{EP}(\kappa)$ that is increasing in κ. If instead, the agent decides on gift-consumption bundles before knowing about her productive type, the associated program is given by:

$$\max_{\{(g_h,x_h),(g_l,x_l)\}} \sum_{n\in\{l,h\}} p_n \left(\theta \cdot \mathcal{G}^{EP}(g_n;\kappa,\bar{g}_h,\bar{g}_l)(g_n;\kappa,\bar{g}_h,\bar{g}_l) + \log x_n\right) \tag{76}$$
$$\text{subject to:} \quad (x_n,g_n) \in \mathcal{B}(x_n,g_n;w_n), \quad \forall n \in \{l,h\}.$$

The solution for the ex ante case in (76) is such that:

1. The equilibrium levels of voluntary contributions and consumption are given by:

$$\hat{g}_n^{EA}(\kappa) = w_n - \frac{1}{\theta\kappa}, \text{ and } \hat{x}_n^{EA}(\kappa) = \frac{1}{\theta\kappa}.$$

2. The equilibrium provision of the public good is given by:

$$\hat{G}^{EA}(\kappa) = W - \frac{1}{\theta\kappa}.$$

When preferences are quasilinear, *Homo moralis* agents that decide behind the veil of ignorance choose the same level of consumption independent of their productivity type. Also, as opposed to the ex post case, contributions are no longer dependent on the population weight p_n. At last, the equilibrium public good provision is higher in the ex ante compared to the ex post case.

In order to compare the equilibria characterized above with a measure of welfare, consider the utilitarian welfare function $\mathcal{W}(G, x_h, x_l) = \theta G + \sum_{n\in\{l,h\}} x_n$. The welfare maximizing bundle $(G^*, x_n^*(w_n))_{n=\{l,h\}}$ solves the program:

$$\begin{aligned} &\max_{\{G,x_l,x_h\}} \mathcal{W}(G, x_h, x_l) \\ &\text{subject to:} \quad \sum_{n\in\{l,h\}} p_n(w_n - x_n) \geq G, \quad \forall n \in \{l, h\}. \end{aligned} \tag{77}$$

The welfare-maximizing solution to (77) is such that:

1. The levels of voluntary contributions and consumption are given by:

$$g_n^* = w_n - \frac{1}{\theta} > \hat{g}_n^{EA}(\kappa) > \hat{g}_n^{EP}(\kappa), \text{ and } x_n^*(\kappa) = \frac{1}{\theta\kappa} < \hat{g}_n^{EA}(\kappa) < \hat{g}_n^{EP}(\kappa).$$

2. The optimal provision of the public good is given by:

$$G^*(\kappa) = W - \frac{1}{\theta} > \hat{G}^{EA}(\kappa) > \hat{G}^{EP}(\kappa).$$

   And the ex ante equilibrium public good provision when$\kappa = 1$ coincides with the welfare maximizing levels of public good provision: $\hat{G}^{EA}(1) = W - \frac{1}{\theta} = G^*(\kappa)$

7.4 Application: State Capacity

Consider a Government ruled by a potentially corrupt elite. The elite may decide to steal a proportion $\rho \in [0, 1]$ of the contribution made by the citizens. Corruption constitutes a public bad, since it erodes democratic institutions, or is linked to illegal activities. A proportion $(1-\rho)$ of all contributions is devoted to the a gross public good H, where $H = (1-\rho)\int_I g_i d\mu(i)$. The remaining proportion ρ constitutes a public bad B, where $B = \rho\beta\int_I g_i d\mu(i)$, and $\beta \in [0, 1]$ is a parameter that measures the damage caused by the appropriation of resources by the elite. If $\beta = 0$, then stealing a fraction ρ of the contributions does not result in any

direct harm to the citizens, besides the effect of the reduction in outcome public good H. In general, for any β, stealing a fraction ρ of the contributions causes direct harm of *beta* to the citizens. The net public good, in this case, is given by:

$$G = H - G = \underbrace{[(1 - (1 + \beta)\rho)]}_{\gamma} \int_I g_i d\mu(i). \tag{78}$$

Consider an entirely corrupt elite that is only concerned about maximizing the resources that it can appropriate by selecting a rate of extraction ρ that raises the highest amount of resources:

$$\max_{\rho} \rho \cdot \hat{G}(\rho; \kappa). \tag{79}$$

The stealing rate that solves this program is given by $\hat{\rho}$:

$$\hat{\rho}(\kappa, \beta) = \frac{1}{1 + \beta} \left(1 - \sqrt{\frac{1}{\kappa \theta W}} \right).$$

The stealing-maximizing rate is increasing in the morality parameter and decreasing in the rate of harm β. Moreover, I show that equilibrium public good provision is increasing in κ, even when the stealing rate also increases in κ. This should be taken as a relevant conclusion limit conclusion: even in the extreme case in which there is an elite whose only interest is to maximize the amount of stolen resources, an economy with Kantian moral agents induces positive provision of public goods.

Kantian morality reciprocity with the state, and tolerance for corruption

The state capacity application presented above yields interesting results in the linear income taxation setting. Assume that the citizen's utility function is given by (16) and that there is a ruling elite that chooses the stealing rate ρ to solve (79) taking the tax rate τ as given. Figure 9 presents the Elite's objective function as a function of the stealing rate ρ for a low (0.1) and a high (0.9) level of β, respectively. From the graph, it is evident that there exists

an interior stealing rate $\hat{\rho}(\kappa)$ that solves (79). Surprisingly, $\hat{\rho}(\kappa)$ is increasing in the degree of morality κ. However, if we compare the plot on the left with the one on the right, we can see how a high β limits the ability of the elite to steal from the public good. The last observation becomes natural if we consider β as a measure of tolerance towards corruption: societies with a low tolerance to corruption (high β) limit the elite's capacity to appropriate resources.

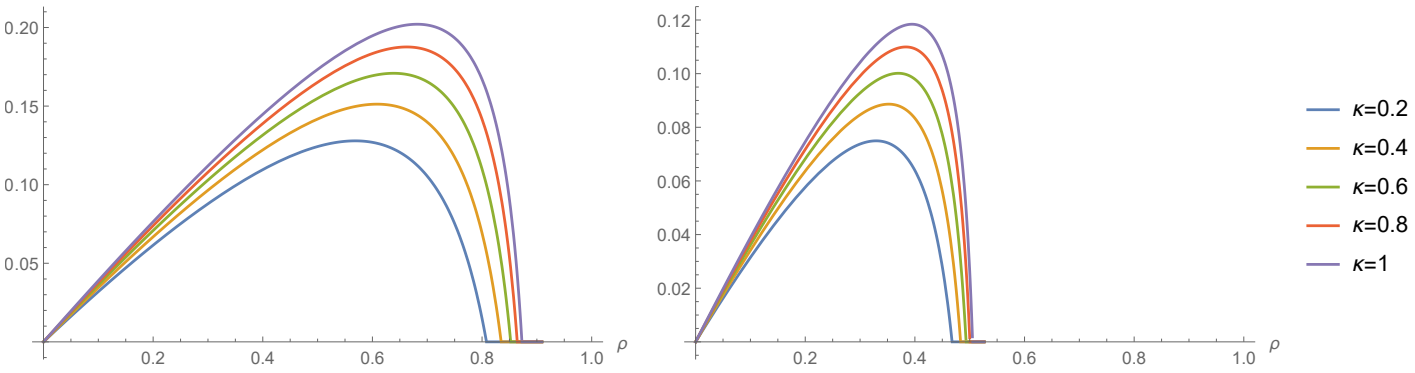

Figure 9: Elite's extraction function for $\tau = 0.5$, $\alpha = 0.5$, $\beta = 0.1$ (left), $\beta = 0,9$ (right).

7.5 A general model of public good provision with moral agents

The present section explores the voluntary contributions equilibria for *Homo moralis* agents with a homogeneous degree of morality. In a simple setting with inelastic labor supply, I show how the degree of morality κ, is the central parameter that pins down the necessary conditions for a voluntary contribution equilibrium both for the case in which types are known by the agents at the moment of deciding upon their contributions (which I refer to as the ex post case) and for the case in which only the distribution of types is known (ex ante case).

For this, consider the private contribution problem: given her (expected) productivity, each agent i chooses the levels of private consumption x_n and *voluntary* contribution to the public good g_n that maximize her utility, for allocations that lie inside of her budget set, taking as given the profile of contributions of the other agents in the continuum. I will compare the equilibrium outcome of this voluntary contribution game to the benchmark first-best result derived in the previous proposition.

The ex ante case. When agents' undertake their contribution decisions "behind the veil of ignorance", they commit to consumption-contribution bundles $\{(x_n, g_n)\}_{i=l,h}$ before knowing their true types. They then select bundles optimally in order to maximize their expected utility. Formally, they solve the following program:

$$\max_{\{(x_l,g_l),(x_h,g_h)\}} \sum_{n\in\{l,h\}} p_n \cdot U(\mathcal{G}(g_l,g_h,\kappa),x_n) \tag{80}$$
$$\text{subject to:} \quad (x_n,g_n) \in \mathcal{B}(x_n,g_n;w_n), \quad \forall i \in \{l,h\}.$$

I would like to draw attention to two features of the program (80). First, the virtual valuation of the public good ($\mathcal{G}(g_l,g_h,\kappa)$) that appears in the problem above depends on the contributions of all the possible types. This feature will be a crucial driver of our result, as agents perceive utility for the global distribution of contributions to the public good instead of just that one of their particular type. Second, the budget set on the constraint above has a simple structure, given by the convex set:

$$(x_n,g_n) \in \mathcal{B}(x_n,g_n,w_n) = \{(x_n,g_n) : x_n + g_n \leq w_n\}, \quad \forall i \in \{l,h\}. \tag{81}$$

The following result solves the optimization problem of the agents and provides a simple necessary condition for any equilibria of voluntary contributions.

Proposition 7 (ex ante optimization)**.** *If agents do not know their own productivities at the moment of optimizing, then a necessary condition for any interior private voluntary provision equilibrium of consumption-contribution bundles* $\{(\hat{x}_l(\tilde{w}),\hat{g}_l(\tilde{w})),(\hat{x}_h(\tilde{w}),\hat{g}_h(\tilde{w}))\}$ *is given by:*

$$\frac{\mathbb{E}_{\tilde{w}} U_2\left(\mathcal{G}(\hat{g}_l(\tilde{w}),\hat{g}_h(\tilde{w}),\kappa),\hat{x}(\tilde{w})\right)}{\mathbb{E}_{\tilde{w}} U_1\left(\mathcal{G}(\hat{g}_l(\tilde{w}),\hat{g}_h(\tilde{w}),\kappa),\hat{x}(\tilde{w})\right)} = \kappa, \tag{82}$$

where $\tilde{w}$ *corresponds to the random productivity that is equal to* w_h *with probability* p *and to* w_l *with probability* $1-p$*.*

Proof. Assume that the inequality in the budget set binds for all agents:

$$x_n + g_n = w_n, \quad \text{for } i \in \{l,h\}. \tag{83}$$

Substitute (83) in the objective function in (80) and obtain:

$$\max_{g_l, g_h} \sum_{n \in \{l,h\}} p_n \cdot U(\mathcal{G}(g_l, g_h, \kappa), w_n - g(w_n)). \tag{84}$$

Consider the necessary first-order conditions for an interior solution in the contribution-space $(g_l, g_h) \in (0, w_l) \times (0, w_h)$. Omitting the arguments inside the moral valuation of the public good $\mathcal{G}(\hat{g}_l(\tilde{w}), \hat{g}_h(\tilde{w}), \kappa)$, they write (for g_h and g_l respectively):

$$p_h \left[-U_2(\mathcal{G}, w_h - \hat{g}_h) + \kappa \cdot p_h \cdot U_1(\mathcal{G}, w_h - \hat{g}_h)\right] + p_l U_1(\mathcal{G}, w_l - \hat{g}_l) \cdot (\kappa \cdot p_h) = 0 \tag{85}$$

$$p_l \left[-U_2(\mathcal{G}, w_l - \hat{g}_l) + \kappa \cdot p_l \cdot U_1(\mathcal{G}, w_l - \hat{g}_l)\right] + p_h U_1(\mathcal{G}, w_l - \hat{g}_l) \cdot (\kappa \cdot p_l) = 0 \tag{86}$$

Solve these equations for the marginal utility of the public good U_2 and obtain:

$$U_2(\mathcal{G}, w_h - \hat{g}_h) = \kappa[p_h U_1(\mathcal{G}, w_h - \hat{g}_h) + p_l U_1(\mathcal{G}, w_l - \hat{g}_l)] \tag{87}$$

$$U_2(\mathcal{G}, w_l - \hat{g}_l) = \kappa[p_h U_1(\mathcal{G}, w_h - \hat{g}_h) + p_l U_1(\mathcal{G}, w_l - \hat{g}_l)] \tag{88}$$

Next, simply multiply by p_h and p_l the previous expressions respectively, and sum the two to obtain:

$$\frac{\mathbb{E}_{\tilde{w}} U_2\left(\mathcal{G}(\hat{g}_l(\tilde{w}), \hat{g}_h(\tilde{w}), \kappa), \hat{x}(\tilde{w})\right)}{\mathbb{E}_{\tilde{w}} U_1\left(\mathcal{G}(\hat{g}_l(\tilde{w}), \hat{g}_h(\tilde{w}), \kappa), \hat{x}(\tilde{w})\right)} = \kappa, \tag{89}$$

□

When agents optimize "behind the veil of ignorance", the ratio of the expected marginal utility of private consumption and expected marginal utility from consumption of the public good must be equal to the degree of morality of the population κ. In any equilibrium, it must be the case that the moral valuation of the public good coincides with the real provision of the public good. This "consistency requirement" writes:

$$\hat{G}(\kappa) = \mathcal{G}(\hat{g}_l(\tilde{w}), \hat{g}_h(\tilde{w}), \kappa), \tag{90}$$

where $\hat{G}$ denotes the equilibrium provision of the public good. The combination of (90) and (82) yields an important takeaway: in any ex ante equilibrium, public good provision $\hat{G}(\kappa)$ increases as the degree of morality of the population increases. In particular, the equilibrium public good provided in an economy populated by materially self-interest agents with $\kappa = 0$ is null. By contrast, in an economy in which types are not observed before contributions are made, Kantian moral agents with $\kappa > 0$ allow for the existence of equilibria with positive provision of the public good. Moreover, since the morality channel enters through the agents' valuation of the public good, increases in the marginal utility of the public good yield increases in equilibrium provision of the public good as long as $\kappa > 0$.

The ex post case. Consider know the case in which agents have knowledge of their productivities before undertaking their decision to contribute to the public good. Any given agent with productivity type w_n for $i \in \{l, h\}$ maximized it's utility for consumption-contribution bundles that lie inside her budget set:

$$\begin{aligned} &\max_{\{x_n, g_n\}} U(\mathcal{G}(g_n, \kappa), x_n) \\ &\text{subject to:} \quad (x_n, g_n) \in \mathcal{B}(x_n, g_n; w_n). \end{aligned} \tag{91}$$

Two important features should be emphasized regarding program (91). First, notice that the term $\mathcal{G}(g_n, \kappa)$ is a function of each agent's contribution, through the moral valuation of the public good. This feature is also present in the "warm glow giving" models (Andreoni (1988), (Andreoni et al., 1998)) that sought to explain empirical giving patterns that were not consistent with the crowding-out theoretical prediction given by models based on selfish or purely altruist agents. Second, as in the ex ante case, the budget set is given equation (81). The following proposition provides necessary conditions for any interior equilibrium as a function of the morality parameter κ.

Proposition 8 (ex post optimization)**.** *Let the bundles $(\hat{x}_n, \hat{g}_n)$ for $i \in \{l, h\}$ be an interior solution to program (91). Then they must meet the following condition:*

$$\frac{U_2\left(\mathcal{G}(\hat{g}_n), \hat{x}_n\right)}{U_1\left(\mathcal{G}(\hat{g}_n), \hat{x}_n\right)} = p_n \cdot \kappa, \quad \forall\, i \in \{l, h\}. \tag{92}$$

Proof. Assume the budget constraint is binding as in equation (83). Now, write the problem

of an arbitrary agent with productivity w_n as:

$$\max_{g_n} U\left(\mathcal{G}(g_n, \kappa), w_n - g_n\right) \tag{93}$$

Omitting the arguments inside the virtual valuation of the public good $\mathcal{G}$, first-order optimality conditions write:

$$-U_2\left(\mathcal{G}, w_n - \hat{g}_n\right) + \kappa p_n \cdot U_1\left(\mathcal{G}, w_n - \hat{g}_n\right) = 0. \tag{94}$$

Rearrange the last equation to obtain:

$$\frac{U_2\left(\mathcal{G}(\hat{g}_n), \hat{x}_n\right)}{U_1\left(\mathcal{G}(\hat{g}_n), \hat{x}_n\right)} = p_n \cdot \kappa. \tag{95}$$

□

This implies that, in any voluntary provision equilibrium in which agents are aware of their types, the marginal rate of substitution between private consumption and the moral valuation over the public good is equal to the degree of morality weighted by the proportion of each agent's type in the population. This result offers an important theoretical prediction based on evolutionary micro-foundations: in a homogeneous population, private contributions to a public good are an increasing function of the degree of morality of the population. As in the ex ante case, notice that the morality channel implies that increases in the material marginal utility perceived from the public good yield increases in the donations made by any given agent, which translate, in equilibrium, to the higher provision of public goods.

The consistency requirement and convergence to the Samuelson Condition. As in the previous case, the public good supplied in equilibrium must be equal to the virtual valuation of the public good evaluated in the equilibrium contributions:

$$\hat{G} = \mathcal{G}(\hat{g}_l, \hat{g}_h, \kappa). \tag{96}$$

Substitute (96) into the necessary condition provided in the previous proposition and sum over types to obtain:

$$\sum_{n\in l,h} \frac{U_2\left(\hat{G},\hat{x}_n\right)}{U_1\left(\hat{G},\hat{x}_n\right)} = \kappa. \tag{97}$$

In equilibrium, the (unweighted) sum of the marginal rates of substitution of public good for private good must exactly equal the degree of morality of the population, κ. It is clear that equation (97) subtly resembles the Samuelson Condition for optimality of public good provision found in equation Proposition 6. In particular, as $\kappa \to 1$ the condition is equivalent to the necessary condition provided by Samuelson (1954) for an efficient provision of public goods when both types have equal weights. This suggests that homogeneous societies populated by *Homo moralis* should exhibit levels of public good provision closer to the welfare-maximizing ones. Moreover, the fact that the sum is unweighed regardless of the value of $p_h \in (0,1)$ has an important implication: in the ex post case, *Homo moralis* agents act in such a way that optimizes their in-type welfare but disregard the consequences of their behavior on the out-types.

Existence and examples. Appendix 7.6 shows that the existence of the voluntary contribution equilibrium can be guaranteed under standard conditions. Moreover, the following example with Cobb Douglas utilities aims to display the main features of the model.

Example 1. Assume that the utility function is given by:

$$U\left(\mathcal{G}(\kappa), x_n\right) = x_n^{1-\alpha}\mathcal{G}(\kappa)^{\alpha} \tag{98}$$

First order conditions imply that:

$$\frac{1-\alpha}{\alpha}\frac{\mathcal{G}(\kappa)}{x_n} = \frac{1-\alpha}{\alpha}\frac{(1-\kappa)G + \kappa p_n g_n}{w_n - g_n} = \kappa p_n \tag{99}$$

Simplifying and solving for g_n:

$$g_n = \begin{cases} 0 & \text{if } \kappa = 0 \\ \max\left\{\alpha w_n - \frac{1-\kappa}{\kappa}\frac{G(1-\alpha)}{\cdot p_n}, 0\right\} & \text{if } \kappa > 0 \end{cases} \tag{100}$$

In any equilibrium in which all types of agents contribute a positive quantity, the equilib-

rium provision public good G^* should satisfy:

$$G^* = \sum_n p_n \cdot g_n(G^*) = \alpha \sum_n p_n w_n - \frac{1-\kappa}{\kappa}(1-\alpha) n G^* \tag{101}$$

Then solve for G^*:

$$G^* = \frac{\alpha \sum_n p_n w_n}{1 + n\frac{1-\kappa}{\kappa}(1-\alpha)} = \frac{\kappa\alpha \sum_n p_n w_n}{\kappa(1-n+\alpha n) + n - \alpha n} \tag{102}$$

Assume that $n = 2$, which implies that:

$$G^* = \frac{\kappa \sum_n p_n w_n}{(2\alpha - 1)\kappa + 2(1-\alpha)} \tag{103}$$

Plugging in this value in the reaction function and obtaining the equilibrium donations:

$$g_n^*(\kappa) = \max\left\{\alpha w_n - \frac{1-\kappa}{\kappa}\frac{(1-\alpha)}{\cdot p_n}\frac{\kappa \sum_n p_n w_n}{(2\alpha-1)\kappa + 2(1-\alpha)}, 0\right\} \tag{104}$$

It is possible to pin down analytically a threshold value of κ that induces both types to donate. Such a threshold is such that both agents contribute when:

$$\kappa > \max_{i=1,2}\left\{\frac{1-2\pi_n}{1+\frac{\pi_n(2\alpha-1)}{1-\alpha}}\right\} = \max_n \bar{\kappa}_n \tag{105}$$

where π_n denotes the relative wealth of each type:

$$\pi_n = \frac{p_n w_n}{p_n w_n + p_{-i} w_{-i}} \tag{106}$$

The total provision of the public good and the ratio between the Pareto Optimal provision and the equilibrium provision in a decentralized donation equilibrium with moral agents is shown in Figure 5. Notice that, the equilibrium provision of the public good may have a kink in the level of κ in which the low-types enter the market and donate positive amounts.

At last, I present a comparison between the ex post and ex ante equilibria. Provided existence, and given a degree of morality, an economy in which agents undertake decisions

behind the veil of ignorance yields higher public good provision. This is formalized in the following proposition.

Proposition 9 (ex post vs. ex ante equilibria)**.** *For a given degree of morality* $\kappa \in (0,1]$ *let the bundles* $(\hat{x}_n^{ep}, \hat{g}_n^{ep})$ *for* $i \in \{l,h\}$ *be an interior solution to the ex post program in* (91)*, and* $(\hat{x}_n^{ea}, \hat{g}_n^{ea})$ *for* $i \in \{l,h\}$ *be an interior solution to the ex ante program in* (80)*. Then:*

$$\hat{G}^{ea}(\kappa) \geq \hat{G}^{ep}(\kappa). \tag{107}$$

Where $\hat{G}^{ea}(\kappa)$ *and* $\hat{G}^{ep}(\kappa)$ *stand for the equilibrium public good provision for the ex ante and ex post case respectively.*

Proof. Included in mathematical appendix 7.7. □

Ex ante decision-making yields higher equilibrium public good provision than the case in which agents learn their productivities before deciding upon contributions. The intuition for the result is the following: when *Homo moralis* agents decide to decrease their contributions behind the veil of ignorance, they internalize the disutility that they would generate on themselves were their type be realized to be the opposite. In this sense, Kantian reasoning becomes amplified by the ex ante optimization.

For the rest of the paper, I focus my attention on the ex post case. The proposition above should serve as proof that all the results obtained below can be interpreted as providing a lower bound on equilibrium public good provision in different environments.

7.6 Existence of the voluntary contribution equilibrium

After characterizing the necessary conditions for the private provision for public goods in an economy populated with moral agents, it is relevant to address the question of when can we expect a voluntary contribution equilibrium to exist, when agents know their types. For this, consider the problem by agent i:

$$\begin{aligned} \max_{x_n, g_n} \quad & U((1-\kappa)G + \kappa p_n g_n, x_n) \\ \text{s.t: } & x_n + g_n \leq w_n \\ & \int_n g_n(w_n) di = G \end{aligned} \tag{108}$$

If we focus on Nash equilibria, agent i takes the equilibrium provision of the public good, G^*, as given. Using this and the fact that in equilibrium the budget constraint must bind we can write the previous problem as:

$$\max_{g_n} \quad U((1-\kappa)G^* + \kappa p_n g_n, w_n - g_n,) \tag{109}$$

We can define the function $S(g_n; w_n, G^*, \kappa)$ as the derivative of U in the last expression with respect to g_n. First Order Conditions guarantee that in a neighbourhood around the optimal g_n (henceforth g_n^*), it is the case that:

$$S(g_n^*; w_n, G^*, \kappa) = -U_1(w_n - g_n^*, (1-\kappa)G^* + \kappa p_n g_n^*) + p_n \kappa \cdot U_2(w_n - g_n^*, (1-\kappa)G^* + \kappa p_n g_n^*) \leq 0 \tag{110}$$

With equality, if the solution is interior. The last equation, together with the fact that second-order conditions guarantee that around any optimal g_n it must be that $\partial S / \partial g_n(g_n^*) \leq 0$. This implies that we can use the Implicit Function Theorem to find the demand function for donating to the public good, which we shall denote $f(w_n, G^*, \kappa)$. Therefore, in an equilibrium it must the case that $g_n^* = f(w_n, G^*, \kappa)$. Hence in any Nash equilibrium, it must be the case that:

1. All the agents solve the program (109), and hence have demand functions for donating given by $g_n^* = f(w_n, G^*, \kappa)$, and second order conditions hold, which means that for a neighbourhood around g_n^i it is the case that $\partial S_n / \partial g_n(g_n^*) < 0$
2. The equilibrium provision of the public good G^* is given by:

$$G^* = \sum_n p_n \cdot f(w_n, G^*, \kappa), \tag{111}$$

i.e, the function $\sum_n p_n \cdot f(w_n, G, \kappa)$ has a fixed point exactly at $G = G^*$.

Theorem 1. *A Nash Equilibrium exists.*

Proof. By the Implicit Function Theorem, we can pin-down the partial derivatives of $f(\cdot)$ with respect to w_n, G^* and κ:

$$\frac{\partial f}{\partial w_n} = \frac{\frac{\partial S}{\partial w_n}}{-\frac{\partial S}{\partial g_n}}, \quad \frac{\partial f}{\partial G^*} = \frac{\frac{\partial S}{\partial G^*}}{-\frac{\partial S}{\partial g_n}}, \quad \frac{\partial f}{\partial \kappa} = \frac{\frac{\partial S}{\partial \kappa}}{-\frac{\partial S}{\partial g_n}} \tag{112}$$

By assumption, second order conditions imply that $-\frac{\partial S}{\partial g_n} > 0$, which means that the sign of all the partial derivatives above is entirely pinned down by the sign of the numerator. Hence:

$$\text{Sign}\left(\frac{\partial f}{\partial w_n}\right) = \text{Sign}\left(-U_{1,1}(\cdot) + \kappa p_n U_{2,1}(\cdot)\right) > 0 \tag{113}$$

$$\text{Sign}\left(\frac{\partial f}{\partial G^*}\right) = \text{Sign}\left(-U_{1,2}(\cdot)(1-\kappa) + \kappa p_n U_{2,2}(\cdot)(1-\kappa)\right) < 0 \tag{114}$$

$$\text{Sign}\left(\frac{\partial f}{\partial \kappa}\right) = \text{Sign}\left(-U_{1,2}(\cdot)(-G^* + p_n \cdot g_n^*) + \kappa p_n U_{2,2}(\cdot)(-G^* + p_n \cdot g_n^*)\right) > 0. \tag{115}$$

The equations described by (113) imply that donations are a normal good, that the demand for donations is decreasing in the total Nash provision of the public good G^*, and that donations are increasing in the degree of morality κ. Now, consider the expression:

$$G^* = \sum_n p_n \cdot \max\{f(w_n, G^*, \kappa), 0\}. \tag{116}$$

Define the set $\hat{W} = \{r \in \mathbb{R}^n : 0 \leq r_n \leq w_n\}$. Notice that for any given κ, $\sum_n p_n \cdot \max\{f(r_n, G, \kappa), 0\}$ maps $\hat{W}$ into itself: hence by Brouwer's Fixed Point Theorem, the must exist a fixed point, this is the Nash Equilibrium vector of donations. □

7.7 Proof of Proposition 9

For $i \in \{l, h\}$, define by $(\hat{x}_n^{ea}, \hat{g}_n^{ea})$ and $(\hat{x}_n^{ep}, \hat{g}_n^{ep})$ the equilibrium private consumption and contribution to the public good for the ex ante and ex post case respectively. Furthermore, for any $\kappa \in (0, 1]$ define by $\hat{G}^{ea}(\kappa)$ and $\hat{G}^{ep}(\kappa)$ the resulting equilibrium public good provision for the two cases.

By contradiction, assume that $\hat{G}^{ea}(\kappa) < \hat{G}^{ep}(\kappa)$. By Proposition 7 and our assumptions on the utility function $U(\cdot)$ we know that:

$$\frac{\sum_{n\in\{l,h\}} p_n \cdot U_2\left(\hat{G}^{ep}, \hat{x}_n^{ea}\right)}{\sum_{n\in\{l,h\}} p_n \cdot U_1\left(\hat{G}^{ep}, \hat{x}_n^{ea}\right)} > \frac{\sum_{n\in\{l,h\}} p_n \cdot U_2\left(\hat{G}^{ea}, \hat{x}_n^{ea}\right)}{\sum_{n\in\{l,h\}} p_n \cdot U_1\left(\hat{G}^{ea}, \hat{x}_n^{ea}\right)} = \kappa, \tag{117}$$

Since $\hat{G}^{ea}(\kappa) < \hat{G}^{ep}(\kappa)$, the consistency requirement in (90) defines two cases to consider. In the first, $\hat{g}_n^{ea} < \hat{g}_n^{ep}$ for all $i \in \{l, h\}$. In the second, $\hat{g}_j^{ea} < \hat{g}_j^{ep}$ for exactly one $j \in \{l, h\}$. We consider both cases below.

1. If $\hat{g}_n^{ea} < \hat{g}_n^{ep}$ for all $i \in \{l, h\}$, then $\hat{x}_n^{ea} > \hat{x}_n^{ep}$ for all $i \in \{l, h\}$. Inequality (117) then implies:

$$\frac{\sum_{n\in\{l,h\}} p_n \cdot U_2\left(\hat{G}^{ep}, \hat{x}_n^{ep}\right)}{\sum_{n\in\{l,h\}} p_n \cdot U_1\left(\hat{G}^{ep}, \hat{x}_n^{ep}\right)} > \kappa. \tag{118}$$

   Applying Proposition 8 to the right side of the inequality in (118) yields:

$$\frac{\sum_{n\in\{l,h\}} p_n \cdot U_2\left(\hat{G}^{ep}, \hat{x}_n^{ep}\right)}{\sum_{n\in\{l,h\}} p_n \cdot U_1\left(\hat{G}^{ep}, \hat{x}_n^{ep}\right)} > \sum_{i\in\{l,h\}} \frac{U_2\left(\hat{G}^{ep}, \hat{x}_n^{ep}\right)}{U_1\left(\hat{G}^{ep}, \hat{x}_n^{ep}\right)} = \kappa. \tag{119}$$

   Without loss of generality, assume that $\hat{x}_h^{ep} > \hat{x}_l^{ep}$. We can apply this to the right-side of inequality (119):

$$\frac{U_2\left(\hat{G}^{ep}, \hat{x}_h^{ep}\right)}{U_1\left(\hat{G}^{ep}, \hat{x}_h^{ep}\right)} > \frac{\sum_{n\in\{l,h\}} p_n \cdot U_2\left(\hat{G}^{ep}, \hat{x}_n^{ep}\right)}{\sum_{n\in\{l,h\}} p_n \cdot U_1\left(\hat{G}^{ep}, \hat{x}_n^{ep}\right)} > \sum_{n\in\{l,h\}} \frac{U_2\left(\hat{G}^{ep}, \hat{x}_n^{ep}\right)}{U_1\left(\hat{G}^{ep}, \hat{x}_n^{ep}\right)}. \tag{120}$$

   Subtracting to the two extremes of the above inequality yields and using the necessary condition in Proposition (8) for the l-types yields:

$$0 > \frac{U_2\left(\hat{G}^{ep}, \hat{x}_l^{ep}\right)}{U_1\left(\hat{G}^{ep}, \hat{x}_l^{ep}\right)} = \kappa \cdot p_l.$$

A clear contradiction for any $\kappa \in (0,1]$.

2. Without loss of generality, assume that $\hat{g}_h^{ea} < \hat{g}_h^{ep}$, but $\hat{g}_l^{ea} > \hat{g}_l^{ep}$. This implies that $\hat{x}_h^{ea} > \hat{x}_h^{ep}$, and $\hat{x}_l^{ea} < \hat{x}_l^{ep}$. We will replicate the argument made for the last case by comparing $\hat{x}_h^{ep}$ and $\hat{x}_l^{ea}$. If $\hat{x}_h^{ep} < \hat{x}_l^{ea}$, then inequality (119) yields $0 > \frac{U_2\left(\hat{G}^{ep},\hat{x}_l^{ep}\right)}{U_1\left(\hat{G}^{ep},\hat{x}_l^{ep}\right)} = \kappa \cdot p_l$, the same contradiction as above. If $\hat{x}_h^{ep} > \hat{x}_l^{ea}$, inequality (119) and the fact that $\hat{x}_h^{ea} > \hat{x}_h^{ep}$ imply $0 > \frac{U_2\left(\hat{G}^{ep},\hat{x}_h^{ep}\right)}{U_1\left(\hat{G}^{ep},\hat{x}_h^{ep}\right)} = \kappa \cdot p_h$, another contradiction.

Finally, it follows that $\hat{G}^{ea}(\kappa) \geq \hat{G}^{ep}(\kappa)$ for any $\kappa \in (0,1]$.

7.8 Proofs for Section 4

-draft- Rewrite the problem in terms of pre-tax income using the equations: $l = y/w$, $x = (1-\tau)y$ and $G = \tau y$. Notice that the program (13) writes then as:

$$\max_{\tau} \theta(\tau y(\tau)) + (1-\tau)y(\tau) - u\left(y(\tau)/w\right) \tag{121}$$

$$\text{subject to: } y(\tau) = w - \frac{1}{1-\tau+\tau\theta\kappa}. \tag{122}$$

The necessary first-order condition to this problem write:

$$y'(\tau)\left[\tau\theta + (1-\tau) - u'(y(\tau)/w)\right] + y(\tau)[\theta - 1] = 0 \tag{123}$$

where $y'(\tau) = \frac{\theta\kappa - 1}{1-\tau+\tau\theta\kappa}$. Rewrite the previous equation as:

$$\left[\tau\theta + (1-\tau) - u'(y(\tau)/w)\right] = \frac{-y(\tau)}{y'(\tau)}[\theta - 1] \tag{124}$$

Now, recall that $u' < 0$. Assume now that $\theta > 1$, this means that the labor supply is increasing in τ and therefore $y'(\tau) > 0$. Hence, the right-hand side of the equation above is decreasing, while the left-hand side is increasing in τ, hence there exists a fixed-point $\tau^*(\kappa;\theta,w)$ equates both sides of the above expression.

7.9 Baseline problem with selfish agents

The baseline problem with selfish agents ($\kappa = 0$). Consider the planner's asymmetric information problem. Since productivities are private information, two self-selection constraints (one for each type of agent) must be satisfied for the tax schedule to be incentive-compatible: I will call them (IC_h) for agents of the high type, and (IC_l) for agents of the low type. These two constraints assure that agents endowed with high (low) productivity do not find it profitable to claim to the government that they have a low (high) productivity. The government's budget constraint (BC) must be met, it ensures that the public good is financed exclusively through the proceeds from the income taxes for both types of agents. Together, these three conditions write:

$$(IC_h): \quad V^h\left(\mathcal{G}(y_h), x_h, y_h\right) \geq V^h\left(\mathcal{G}(y_l), x_l, y_l\right) \tag{125}$$

$$(IC_l): \quad V^l\left(\mathcal{G}(y_l), x_l, y_l\right) \geq V^l\left(\mathcal{G}(y_h), x_h, y_h\right) \tag{126}$$

$$(BC): \quad G = p \cdot \tau(y_h) + p_l \cdot \tau(y_l) \tag{127}$$

Proposition 10. *When $\kappa = 0$, and denoting $\lambda_h > 0$ as the Lagrangian multiplier corresponding to IC_h, the solution to the problem defined in (35) is such that:*

6.1 ***No distortion at the top.*** *The marginal tax rate for the able agents is equal to zero:*

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

6.2 ***Distortion at the bottom.*** *The marginal tax rate for the less able agents is positive:*

$$\psi_l(G, w_l, y_l) < 1;$$

Proof. Included in mathematical appendix. □

The Lagrangian associated with problem (35) writes:

$$\mathcal{L}\left(x_h, x_l, y_h, y_l, G\right) = p \cdot V^h\left(G, x_h, y_h\right) + (1-p) \cdot V^l\left(G, x_l, y_l\right) + \lambda_h \left(V^h\left(G, x_h, y_h\right) - V^h\left(G, x_l, y_l\right)\right) \tag{128}$$

$$+ \lambda_l \left(V^l\left(G, x_l, y_l\right) - V^l\left(G, x_h, y_h\right)\right) + \mu\left(p \cdot (y_h - x_h) + (1-p) \cdot (y_l - x_l) - G\right) \tag{129}$$

The necessary first-order conditions to this problem write:

$$\frac{\partial \mathcal{L}}{\partial x_h} = p \cdot V_2^h\left(\mathcal{G}, x_h, y_h\right) + \lambda_h \cdot V_2^h\left(\mathcal{G}, x_h, y_h\right) - \lambda_l \cdot V_2^l\left(\mathcal{G}, x_h, y_h\right) - \mu \cdot p = 0 \tag{130}$$

$$\frac{\partial \mathcal{L}}{\partial y_h} = p \cdot V_3^h\left(\mathcal{G}, x_h, y_h\right) + \lambda_h \cdot V_3^h\left(\mathcal{G}, x_h, y_h\right) - \lambda_l \cdot V_3^l\left(\mathcal{G}, x_h, y_h\right) + \mu \cdot p = 0 \tag{131}$$

$$\frac{\partial \mathcal{L}}{\partial x_l} = (1-p) \cdot V_2^l\left(\mathcal{G}, x_l, y_l\right) - \lambda_h \cdot V_2^h\left(\mathcal{G}, x_l, y_l\right) + \lambda_l \cdot V_2^l\left(\mathcal{G}, x_l, y_l\right) - \mu \cdot (1-p) = 0 \tag{132}$$

$$\frac{\partial \mathcal{L}}{\partial y_l} = (1-p) \cdot V_3^l\left(\mathcal{G}, x_l, y_l\right) - \lambda_h \cdot V_3^h\left(\mathcal{G}, x_l, y_l\right) + \lambda_l \cdot V_3^l\left(\mathcal{G}, x_l, y_l\right) + \mu \cdot (1-p) = 0 \tag{133}$$

$$\frac{\partial \mathcal{L}}{\partial G} = p \cdot V_1^h\left(\mathcal{G}, x_h, y_h\right) + (1-p) \cdot V_1^l\left(\mathcal{G}, x_l, y_l\right) + \lambda_h \left(V_1^h\left(\mathcal{G}, x_h, y_h\right) - V_1^h\left(\mathcal{G}, x_l, y_l\right)\right) \tag{134}$$

$$+ \lambda_l \left(V_1^l\left(\mathcal{G}, x_l, y_l\right) - V_1^l\left(\mathcal{G}, x_h, y_h\right)\right) - \mu = 0 \tag{135}$$

I focus in the "normal" case in which no low type agent wants to imitate a high type, i.e: assume that (IC_l) is always satisfied ($\lambda_l = 0$) and (IC_h) is binding ($\lambda_h > 0$). Hence, summing up the first two equations, we obtain the no distortion at the top result:

$$\psi_h(\mathcal{G}, x_h, y_h) = \frac{-V_3^h\left(\mathcal{G}, x_h, y_h\right)}{V_2^h\left(\mathcal{G}, x_h, y_h\right)} = 1 \tag{136}$$

This means that the high productivity agents' marginal income tax is equal to zero. On the other hand, we can divide the fourth equation by the the third one and obtain:

$$\psi_l(\mathcal{G}, x_l, y_l) = \frac{-V_3^l(\mathcal{G}, x_l, y_l)}{V_2^l(\mathcal{G}, x_l, y_l)} = \frac{\mu \cdot (1-p) - \lambda_h \cdot V_3^h(\mathcal{G}, x_l, y_l)}{\lambda_h \cdot V_2^h(\mathcal{G}, x_l, y_l) + \mu \cdot (1-p)} \tag{137}$$

$$= \frac{1 - \lambda_h \cdot V_3^h(\mathcal{G}, x_l, y_l)\mu^{-1}(1-p)^{-1}}{1 + \lambda_h \cdot V_2^h(\mathcal{G}, x_l, y_l)\mu^{-1}(1-p)^{-1}} \tag{138}$$

$$= \frac{1 - \lambda_h \cdot \mu^{-1}(1-p)^{-1}V_2^h(\mathcal{G}, x_l, y_l)\psi_h(\mathcal{G}, x_l, y_l)}{1 + \lambda_h \cdot V_2^h(\mathcal{G}, x_l, y_l)\mu^{-1}(1-p)^{-1}} \tag{139}$$

$$= \frac{1 - \lambda_h \cdot \mu^{-1}(1-p)^{-1}V_2^h(\mathcal{G}, x_l, y_l)\psi_h(\mathcal{G}, x_l, y_l) + \psi_h(\mathcal{G}, x_l, y_l) - \psi_h(G, x_l, y_l)}{1 + \lambda_h \cdot V_2^h(\mathcal{G}, x_l, y_l)\mu^{-1}(1-p)^{-1}} \tag{140}$$

$$= \psi_h(\mathcal{G}, x_l, y_l) + \frac{1 - \psi_h(\mathcal{G}, x_l, y_l)}{1 + \lambda_h \cdot V_1^h(\mathcal{G}, x_l, y_l)\mu^{-1}(1-p)^{-1}} \tag{141}$$

$$< 1 \tag{142}$$

where the last inequality follows from the fact that from Assumption 23:

$$\psi_h(\mathcal{G}, x_l, y_l) < \psi_l(\mathcal{G}, x_l, y_l) \tag{143}$$

this means that $\tau'(g_l) > 0$: the marginal tax rate for the low-productivity agents is positive.

Finally, the last equation gives the optimal provision of public good. We can combine the first and third first order conditions to solve for

$$\mu = \mathbb{E}_w \left[V_2^I(\mathcal{G}, x, y)\right] + \lambda_h \left(V_2^h(\mathcal{G}, x_h, y_h) - U_2^h(\mathcal{G}, x_h, y_h)\right) \tag{144}$$

and then substitute in the last first order conditions to obtain:

$$\frac{\mathbb{E}_w \left[V_2^I(\mathcal{G}, x, y)\right]}{\mathbb{E}_w \left[V_1^I(\mathcal{G}, x, y)\right]} = 1 + \lambda_h \left(\frac{V_1^h(\mathcal{G}, x_h, y_h) - V_1^h(\mathcal{G}, x_l, y_l)}{\mathbb{E}_w \left[V_1^I(\mathcal{G}, x, y)\right]} - \frac{V_2^h(\mathcal{G}, x_h, y_h) - V_2^h(\mathcal{G}, x_l, y_l)}{\mathbb{E}_w \left[V_1^I(\mathcal{G}, x, y)\right]} \right) \tag{145}$$

note that when $\lambda_h = 0$ we are in the first-best solution. Whenever $\lambda_h > 0$, however, the equilibrium provision of public good is reduced provided that the term in brackets is negative.

7.10 Proof of Proposition 2

The problem faced by an agent of type i is given by $\max_{x_n,l_n}$ subject to $x_n \leq y_n - \tau(y_n)$ and $y_n = w_n l_n$.In an interior equilibrium the budget constraint binds and it is possible to write the agent's problem as a function of pre-tax income y_n as follows:

$$\max_{y_n} V^i\left(\mathcal{G}(y), y_n - \tau(y_n), y\right) \tag{146}$$

where $\mathcal{G}$ is defined according to (34). For each type of agent, the necessary first order conditions to this problem write:

$$V_2^h\left(\mathcal{G}, x_h, y_h\right)\left(1 - \tau'(y_h)\right) + \kappa p_h \tau'(y_h) \cdot V_1^h\left(\mathcal{G}, x_h, y_h\right) + V_3^h\left(\mathcal{G}, x_h, y_h\right) = 0 \tag{147}$$

$$V_2^l\left(\mathcal{G}, x_l, y_l\right)\left(1 - \tau'(y_l)\right) + \kappa p_l \tau'(y_l) \cdot V_1^l\left(\mathcal{G}, x_l, y_l\right) + V_3^l\left(\mathcal{G}, x_l, y_l\right) = 0. \tag{148}$$

Dividing the two previous equations by the marginal utility derived from consumption of the private good V_1 we obtain our desired result:

$$\tau'(g_h)\left(1 + p \cdot \kappa \cdot \phi(\mathcal{G}, w_h)\right) = 1 - \psi(\mathcal{G}, w_h) \tag{149}$$

$$\tau'(g_l)\left(1 + p_l \cdot \kappa \cdot \phi(\mathcal{G}, w_l)\right) = 1 - \psi(\mathcal{G}, w_l) \tag{150}$$

Since I work with a finite number of types of atomless agents, the optimal tax function $\tau(y)$ may not be differentiable in $y_{i\in\{l,h\}}$ [20]. Hence, solve the last equation for τ' and pin down the marginal tax rate $\tau'(y_n)$:

$$\tau'_\kappa(y_n) \triangleq \frac{1 - \psi(G, w_n)}{1 + p_n \cdot \kappa \cdot \phi(\mathcal{G}, w_n)}. \tag{151}$$

Solution to the quasilinear case

In this subsection, I develop the case in which utilities are quasilinear. The purpose of this is to provide a good illustration of the derivations provided above while revealing the main mechanism that arises in the general solution to the program (35). Consider then the following example.

Example 2 (The full problem with quasilinear utilities)**.** Assume that agents have utilities of the form:

[20] As argued by Stiglitz (1982), however, there exists tax structures for which the right-hand-side of equation (151) is the left-hand derivative of the tax schedule τ at $y_n = w_n l_n$.

$$V^j(\mathcal{G}(y_j), x_j, y_j) = A^j(x_j, y_j) + \theta \cdot \mathcal{G}(\kappa; y_j), \quad \text{for } \theta \geq 1. \tag{152}$$

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:

$$\psi_l(w_l) = \frac{-\partial A^h(x_l, y_l)/\partial y_l}{\partial A^h(x_l, y_l)/\partial x_l} < \frac{-\partial A^l(x_l, y_l)/\partial y_l}{\partial A^l(x_l, y_l)/\partial x_l} = \psi_l(w_h). \tag{153}$$

From the previous equation, notice that quasi linearity implies that single crossing is independent from the consumption of the public good. On the other hand, the individual rationality (decentralization) result allows us to pin-down the marginal tax rate for any type $j \in \{l, h\}$ as:

$$\tau'(y_j) = \frac{1 - \psi_j(w_j)}{1 - \theta \cdot \kappa \cdot p_j / A^j_{x_j}}, \tag{154}$$

where in general $A^i_{x_j} = \partial A^i(x_j, y_j)/\partial x_j$ (in particular, $i = j$ in the equation above). Recurring to the Revelation Principle we can then write V^j as a function of x and y only. In this case, using the definition of the moral valuation of the public good presented above:

$$V^j(\mathcal{G}(y_n), x_n, y_n) = A^j(x_n, y_n) + \theta \cdot \left[(1 - \kappa) \cdot G + \kappa \cdot p_j \cdot (y_j - x_j)\right]. \tag{155}$$

Equation (155) allows us to write the incentive constraints of the high-ability agents as:

$$A^h(x_h, y_h) - A^h(x_l, y_l) \geq \kappa \cdot p_h \cdot \theta \left[(y_l - x_l) - (y_h - x_h)\right]. \tag{156}$$

This provides a clear intuition regarding the effect of the morality parameter κ over the incentive constraint: Kantian morality relaxes the incentive constraint for high-ability types if and only if $y_h - x_h > y_l - x_l$. This is because high-ability agents are motivated by higher average income taxes (recall that $\tau(y_h) = y_h - x_h$). Crucially, this means that raising taxes from low-ability agents makes the incentive constraint harder to be met concerning a baseline in which $\kappa = 0$. To see this, start with $\kappa = 0$ in the inequality above, then slightly increase κ:

it is clear that if $y_h - x_h < y_l - x_l$ then the inequality becomes harder to meet for the principal.

Now, consider 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):

$$\begin{aligned}
&\max_{x_h,x_l,y_h,y_l} \theta \cdot G + p_h \cdot V^h(x_h, y_h) + p_l \cdot V^l(x_l, y_l) \\
&(BC): \quad p_h \cdot (y_h - x_h) + p_l \cdot (y_l - x_l) \geq G \\
&(IC_h): \quad A^h(x_h, y_h) - A^h(x_l, y_l) \geq -\kappa \cdot p_h \cdot \theta\left((y_h - x_h) - (y_l - x_l)\right) \\
&(IC_l): \quad A^l(x_l, y_l) - A^l(x_h, y_h) \geq -\kappa \cdot p_l \cdot \theta\left((y_l - x_l) - (y_h - x_h)\right)
\end{aligned} \tag{157}$$

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:

$$\begin{aligned}
\mathcal{L}(x_h, y_h, x_l, y_l, \lambda_h) =& \theta \cdot \left(\sum_{j \in l,h} p_j(y_j - x_j)\right) + p_h \cdot V^h(x_h, y_h) + p_l \cdot V^l(x_l, y_l) \\
&+ \lambda_h \left(A^h(x_h, y_h) - A^h(x_l, y_l) + \kappa \cdot p_h \cdot \theta\left((y_h - x_h) - (y_l - x_l)\right)\right)
\end{aligned} \tag{158}$$

The first-order optimality conditions to this problem write:

$$\frac{\partial \mathcal{L}(x_h, y_h, x_l, y_l, \lambda_h)}{\partial x_h} = -\theta \cdot p_h + p_h \cdot A^h_{x_h} + \lambda_h \left(A^h_{x_h} - \theta \cdot \kappa \cdot p_h\right) = 0 \tag{159}$$

$$\frac{\partial \mathcal{L}(x_h, y_h, x_l, y_l, \lambda_h)}{\partial y_h} = \theta \cdot p_h + p_h \cdot A^h_{y_h} + \lambda_h \left(A^h_{y_h} + \theta \cdot \kappa \cdot p_h\right) = 0 \tag{160}$$

$$\frac{\partial \mathcal{L}(x_h, y_h, x_l, y_l, \lambda_h)}{\partial x_l} = -\theta \cdot p_l + p_l \cdot A^l_{x_l} + \lambda_h \left(-A^h_{x_l} + \theta \cdot \kappa \cdot p_h\right) = 0 \tag{161}$$

$$\frac{\partial \mathcal{L}(x_h, y_h, x_l, y_l, \lambda_h)}{\partial y_l} = \theta \cdot p_l + p_l \cdot A^l_{y_l} + \lambda_h \left(-A^h_{y_l} - \theta \cdot \kappa \cdot p_h\right) = 0 \tag{162}$$

$$\tag{163}$$

The above system allows us to characterize completely the solution to the planner's prob-

lem. 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 2) 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:

$$\psi_l(w_l) \triangleq \frac{-A^l_{y_l}}{A^l_{x_l}} = \frac{\theta \cdot p_l - \lambda \left(A^h_{y_l} + \theta \cdot \kappa \cdot p_h\right)}{\theta \cdot p_l + \lambda \left(A^h_{x_l} + \theta \cdot \kappa \cdot p_h\right)} \tag{164}$$

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

$$v = \frac{\lambda_h A^h_{x_l}}{\theta p_l}, \quad \text{and} \quad K(\kappa) = \theta \cdot \kappa \cdot p_h. \tag{165}$$

We can now rewrite the previous equation as:

$$\psi_l(w_l) \triangleq \frac{-A^l_{y_l}}{A^l_{x_l}} = \frac{1 - v \cdot K(\kappa) + \psi_l(w_h)}{1 + v - vK(\kappa)} \tag{166}$$

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

$$\left(1 - v \cdot K(\kappa)\right)\left(1 - \psi_l(w_l)\right) = v \cdot \left(\psi_l(w_h) - \psi_l(w_h)\right) \tag{167}$$

Recall that the single-crossing assumption (153) asserts that $\psi_l(w_h) - \psi_l(w_h) > 0$. This means that, since $1 - v \cdot K(\kappa) > 0$ for any $\kappa \in [0, 1]$, it must be the case that $\psi_l(w_l) < 1$. Hence, the bundle offered by the principal to the less able type is distorted in order to meet the desired self-selection constraints.

Proposition 11. *The marginal tax rate implied by the solution of the planner's program in (157) is such that there exists a threshold value of κ, defined as $\hat{\kappa}(\theta, \lambda_h, p_h, A^l_{x_l}, A^h_{x_l})$, such that: if $\kappa > \hat{\kappa}$ then the marginal tax rate faced by the low-ability type is decreasing in κ, while if $\kappa < \hat{\kappa}$ then the marginal tax rate is increasing in the morality parameter κ. Moreover, it is the case that:*

1. *$\hat{\kappa}(\theta, \lambda_h, p_h, A^l_{x_l}, A^l_{x_l})$ is increasing in p_l and $A^l_{x_l}$;*
2. *$\hat{\kappa}(\theta, \lambda_h, p_h, A^l_{x_l}, A^l_{x_l})$ is decreasing in λ_h, θ, p_h, and $A^h_{x_l}$.*

Proof. From the derivation to the previous problem and the solution to the decentralized problem:

$$\tau'(y_l) = \frac{v(\psi_l(w_l) - \psi_l(w_h))}{\left(1 - \theta\kappa p_l / A^l_{x_l}\right)\left(1 - \lambda A^h_{x_l}\kappa p_h / p_l\right)}.$$

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:

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

□

Figure 10 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 (156)): I call this, the "moral incentive effect"

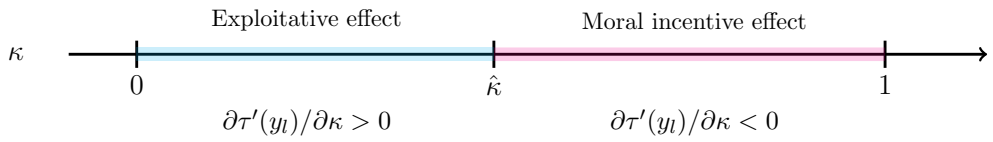

Figure 10: Morality parameter and marginal tax rate of low-ability types.

7.11 Optimal income taxation when the IC of the low-types binds

Proposition 12 (The IC of the low types binds)**.** *When $\kappa \in (0,1]$ and the incentive constraint of the low type is binding, the solution to the problem defined in (35) is such that:*

10.1 ***No distortion at the bottom.*** *the marginal tax payed by the low ability types is equal to zero:*

$$\psi_l(\mathcal{G}, w_l) = 1;$$

10.2 ***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:*

$$\psi_l(\mathcal{G}, w_h) = \gamma(\kappa) > 1, \quad \text{for } \gamma(\kappa) < 0, \text{ and } \gamma'(\kappa) < 0.;$$

Proof. The Lagrangian associated with problem (35) writes:

$$\begin{aligned}\mathcal{L}\left(x_h, y_h, x_l, y_l, G\right) =& p_h \cdot V^h\left(G, x_h, y_h\right) + p_l \cdot V^l\left(G, x_l, y_l\right) + \lambda_l \left(V^l\left(\mathcal{G}, x_l, y_l\right) - V^l\left(\mathcal{G}, x_h, y_h\right)\right) \\ &+ \mu\left(p_h \cdot \left(y_h - x_h\right) + p_l \cdot \left(y_l - x_l\right) - G\right) \end{aligned} \tag{168}$$

The necessary first order conditions to this problem write:

$$\frac{\partial \mathcal{L}}{\partial x_h} = p_h \cdot V_2^h\left(\mathcal{G}, x_h, y_h\right) + \lambda_l \left(-V_2^l\left(\mathcal{G}, x_l, y_l\right) + \kappa \cdot p_l V_1^l\left(\mathcal{G}, x_h, y_h\right)\right) - \mu \cdot p_h = 0 \tag{169}$$

$$\frac{\partial \mathcal{L}}{\partial y_h} = p_h \cdot V_3^h\left(\mathcal{G}, x_h, y_h\right) + \lambda_l \left(-V_3^l\left(\mathcal{G}, x_l, y_l\right) - \kappa \cdot p_l V_1^l\left(\mathcal{G}, x_h, y_h\right)\right) + \mu \cdot p_h = 0 \tag{170}$$

$$\frac{\partial \mathcal{L}}{\partial x_l} = p_l \cdot V_1^2\left(\mathcal{G}, x_l, y_l\right) + \lambda_l \left(V_2^l\left(\mathcal{G}, x_l, y_l\right) - \kappa p_l V_1^l\left(\mathcal{G}, x_l, y_l\right)\right) - \mu \cdot p_l = 0 \tag{171}$$

$$\frac{\partial \mathcal{L}}{\partial y_l} = p_l \cdot V_3^2\left(\mathcal{G}, x_l, y_l\right) + \lambda_l \left(V_3^l\left(\mathcal{G}, x_l, y_l\right) + \kappa p_l V_1^l(x_l, \mathcal{G}, y_l)\right) + \mu \cdot p_l = 0 \tag{172}$$

$$\frac{\partial \mathcal{L}}{\partial \mathcal{G}} = p_h \cdot V_1^h\left(\mathcal{G}, x_h, y_h\right) + p_l \cdot V_1^l\left(\mathcal{G}, x_l, y_l\right) + \lambda_l \left(V_1^h\left(\mathcal{G}, x_h, y_h\right) - V_1^h\left(\mathcal{G}, x_l, y_l\right)\right) - \mu = 0 \tag{173}$$

in the same manner as in the previous proof, summing up the third and fourth equations and using the fact that in equilibrium, the virtual valuation of the public good coincides with the real provision of the public good ($G = \mathcal{G}$), we obtain the no distortion at the bottom result:

$$\psi_l(\mathcal{G}, x_l, y_l) = \frac{-V_3^l(\mathcal{G}, x_l, y_l)}{V_2^l(\mathcal{G}, x_l, y_l)} = 1 \tag{174}$$

This means that the low productivity agents' marginal income tax is equal to zero. On the other hand, we can define again $C(\kappa) = -\kappa V_1^l(x_h, \mathcal{G}, y_h)$, divide the second equation by the the first one and obtain:

$$\frac{V_3^h(x_h, \mathcal{G}, y_h)}{V_2^h(\mathcal{G}, x_h, y_h)} = \frac{-\mu \cdot p_h + \lambda_l \left(V_3^l(\mathcal{G}, x_h, y_h) - C(\kappa)\right)}{\lambda_l \left(V_2^l(\mathcal{G}, x_h, y_h) + C(\kappa)\right) + \mu \cdot p_h} \tag{175}$$

Following the same logic of the previous proof, we can now multiply both sides by $(\lambda_l \left(V_(^l\mathcal{G}, x_h, y_h) + C(\kappa)\right)$ - and obtain:

$$\tau'(y_h)\left(1 + p_h \cdot \kappa \cdot \phi_h(\mathcal{G}, x_h, y_h)\right) \triangleq (1 - \psi_h(\mathcal{G}, x_h, y_h)) = \frac{\lambda_l V_2^l(\mathcal{G}, x_h, y_h)}{\mu \cdot p_h + \lambda_l \cdot C(\kappa)} \left(\psi_h\left(x_h, \mathcal{G}, y_h\right) - \psi_l(\mathcal{G}, x_h, y_h)\right) < 0 \tag{176}$$

Recall that by assumption $\left(\psi_h\left(x_h, \mathcal{G}, y_h\right) - \psi_l(\mathcal{G}, x_h, y_h)\right) < 0$, which yields the desired result. □

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.

8 An application: optimal tax treatment of private contributions

An interesting application of the model presented in the previous section is given by the study of tax-favored voluntary donations. In this line of research, Diamond (2006) presents

a model of optimal income taxation in which agents (which may or may not have warm-glow preferences) can make government-subsidized donations. Such donations can be welfare improving because of two channels: first, higher donations from high-income earners can serve as an instrument that relaxes incentive compatibility constraints for donors. Second, private donations also reduce overall consumption, which relaxes the resource constraint of the economy.

The model proposed by Diamond (2006) features a finite number of agents that belong to each income type, and it predicts that donations will decrease as the number of agents of each type increases. Therefore, unselfish preferences constitute a useful tool that can potentially explain voluntary contributions to public goods in the presence of optimal income taxation.

To keep the exposition simple, consider the following linear additive specification of the agents' utility function:

$$V(x, l, G) = C(x) - a \cdot l + B(G) = C(x) - \left(\frac{a}{w}\right) y + B(G) \tag{177}$$

Notice that the single crossing holds since $a/w_h < a/w_l$. We can further define income before taxes as: $z_n = (1 - \tau(y_n))y_n$. The main framework from the previous section remains unchanged, however, suppose that now the government can subsidize private donations from agents to the public good. In particular, suppose that the government proposes a subsidy i to each type of agent, implies that private consumption is given by: $x_n = z_n - (1 - s_n)g_n$. We can replace this expression into the utility function and obtain, for Homo moralis agents with a degree of morality κ:

$$C(z_n - (1 - s_n)g_n) - \frac{a}{w_n} \cdot l_n + B(\mathcal{G}(\kappa)) \tag{178}$$

Next, we can maximize the previous expression to pin down the optimal gift:

$$(1 - s_n) \cdot C'(x_n) = \kappa p_n B'(\mathcal{G}(\kappa)) \tag{179}$$

Assume that the equilibrium level of public good provision is given by G^*. We can then use the consistency condition $\mathcal{G} = G$ in the above equation to retrieve the optimal level of consumption and private donations: x_n^* and g_n^*.

If private donations are not allowed, the problem solved by the government is given by:

$$\max_{x,G,y} \sum_n p_n \left(C(x_n) - \frac{a}{w_n} \cdot y_n + B(G) \right) \tag{180}$$

s.t: (181)

$$\sum_n p_n (y_n - x_n) \geq G \tag{182}$$

$$C(x_h) - \frac{a}{w_h} \cdot y_h + B(G) \geq C(x_l) - \frac{a}{w_h} \cdot y_l + B(G) \tag{183}$$

As was previously shown, the solution to this problem is such that $C'(x_h^*) = \frac{a}{w_h}$ and $C'(x_h^*) > \frac{a}{w_h}$, which means that the consumption of the low type is decreased with respect to the first-best level.

Now, suppose that the government can now the underlying optimal level G, and has the power to propose optimal gift schedules (which essentially amount to choosing optimal subsidy rate), we then have that the program of the government is given by:

$$\max_{x,G,y,g} \sum_n p_n \left(C(x_n) - \frac{a}{w_n} \cdot y_n + B(G) \right) \tag{184}$$

s.t:

$$\sum_n p_n (y_n - x_n) \geq G$$

$$C(x_h) - \frac{a}{w_h} \cdot y_h + B((1-\kappa)G + \kappa p_h \cdot g_h) \geq C(x_l) - \frac{a}{w_h} \cdot y_l + B((1-\kappa)G + \kappa \cdot p_h \cdot g_l)$$

Notice that moral concerns relax the incentive compatibility constraint, which potentially increases welfare with respect to the baseline scenario in which agents cannot make private contributions.

Example 3. Assume now that $C(x_n) = \sqrt{x_n}$, and $B(G) = G^\gamma$. The first order condition with respect to y_h should yields the optimal level of public good:

$$B'(G^*) = a \left(\frac{1}{w_h} + \frac{1}{w_l} \right) \tag{185}$$

The last condition. together with the no distortion at the top result imply:

$$G^* = \left(\gamma \cdot \frac{w_h \cdot w_l}{a} \right)^{\frac{1}{1-\gamma}}, \quad x_h^* = \frac{1}{4} \left(\frac{w_h}{a} \right)^2 \tag{186}$$

These two conditions together imply that:

$$B'(G^*) = \gamma \left(G^*\right)^{\gamma-1} = B'(G^*) = \gamma \left(\left(\gamma \cdot \frac{w_h \cdot w_l}{a} \right)^{\frac{1}{1-\gamma}} \right)^{\gamma-1} = \frac{a}{w_h \cdot w_l} \tag{187}$$

We can now plug this in the FOC for x_l^*:

$$C'(x_l^*) = \frac{\frac{a}{w_h \cdot w_l} \cdot p_l}{a/w_h - p_h \cdot \frac{a}{w_h \cdot w_l}} \frac{w_h}{a} \tag{188}$$

$$= \frac{\frac{1}{w_l} \cdot p_l}{a/w_h \cdot w_l \left(w_l - p_h\right)} \tag{189}$$

$$= \frac{w_h}{a} \frac{p_l}{w_l + p_l - 1} \tag{190}$$

Then we have that:

$$x_l^* = \frac{1}{4} \left(\frac{a}{w_h} \frac{w_l + p_l - 1}{p_l} \right)^2 \tag{191}$$

Now, we go back to the $2\times$ 2 system that formed by the IC and budget constraint of the government:

$$(IC): \; y_h - y_l = \frac{w_h}{a} \left(C(x_h^*) - C(x_l^*) \right) \tag{192}$$

$$(BC): \; p_h y_h + p_l y_l = G^* - p_h x_h^* - p_l x_l^* \tag{193}$$

We can solve this system and obtain the following solutions for optimal values of pretax income:

$$y_h^* = G^* - \sum_n p_n x_n^* + p_l \frac{w_h}{a} \left(C(x_h^*) - C(x_l^*) \right) \tag{194}$$

$$y_l^* = G^* - \sum_n p_n x_n^* - p_h \frac{w_h}{a} \left(C(x_h^*) - C(x_l^*) \right) \tag{195}$$

Notice that pretax income can be essentially decomposed into two parts: the optimal share of the public good net of consumption plus a distortion term that stems from the IC constraints. Provided monotonicity, it is clear that these high-income agents should enjoy a higher pretax income than low type agents.

References

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

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

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

Alger, I. and Weibull, J. W. (2013). Homo moralis—preference evolution under incomplete information and assortative matching. *Econometrica*, 81(6):2269–2302.

Alger, I. and Weibull, J. W. (2016). Evolution and kantian morality. *Games and Economic Behavior*, 98:56 – 67.

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, 1(3):323 – 338.

Alm, J., Jackson, B. R., and McKee, M. (1993). Fiscal exchange, collective decision institutions, and tax compliance. *Journal of Economic Behavior and Organization*, 22(3):285 – 303.

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

Andreoni, J. (1988). Privately provided public goods in a large economy: The limits of altruism. *Journal of Public Economics*, 35(1):57 – 73.

Andreoni, J. (1990a). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, 100(401):464–477.

Andreoni, J. (1990b). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, 100(401):464–477.

Andreoni, J., Erard, B., and Feinstein, J. (1998). Tax compliance. *Journal of Economic Literature*, 36(2):818–860.

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

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

Bénabou, R. and Tirole, J. (2006). Belief in a Just World and Redistributive Politics*. *The Quarterly Journal of Economics*, 121(2):699–746.

Boadway, R. and Keen, M. (1993). Public Goods, Self-Selection and Optimal Income Taxation. *International Economic Review*, 34(3):463–478.

Bordignon, M. (1993). A Fairness Approach to Income-Tax Evasion. *Journal of Public Economics*, 52(3):345–362.

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

DeBacker, J. M., Heim, B. T., and Tran, A. (2012). Importing corruption culture from overseas: Evidence from corporate tax evasion in the united states. Working Paper 17770, National Bureau of Economic Research.

Diamond, P. (2006). Optimal tax treatment of private contributions for public goods with and without warm glow preferences. *Journal of Public Economics*, 90(4-5):897–919.

Dwenger, N., Kleven, H., Rasul, I., and Rincke, J. (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in germany. *American Economic Journal: Economic Policy*, 8(3):203–32.

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

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

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

Feld, L. P. and Frey, B. S. (2002). Trust breeds trust: How taxpayers are treated. *Economics of Governance*, 3(2):87–99.

Gordon, J. P. P. (1989). Individual morality and reputation costs as deterrents to tax evasion. *European Economic Review*, 33(4):797–805.

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

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

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

Johansen, L. (1977). The theory of public goods: Misplaced emphasis? *Journal of Public Economics*, 7(1):147–152.

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

Klosko, G. (2004). *The principle of fairness and political obligation*. Rowman & Littlefield.

Kountouris, Y. and Remoundou, K. (2013). Is there a cultural component in tax morale? evidence from immigrants in europe. *Journal of Economic Behavior and Organization*, 96(C):104–119.

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

Levi, M. (1988). *Of Rule and Revenue*. University of California Press.

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

Locke, J. (1690). *Locke: Two treatises of government*. Cambridge university press.

Luttmer, E. F. P. and Singhal, M. (2014). Tax morale. *Journal of Economic Perspectives*, 28(4):149–68.

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

Myles, G. (1995). Public Economics. In *Public Economics*, number October, pages 340–440.

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

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

Samuelson, P. A. (1954). The pure theory of public expenditure.

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

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

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

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

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

Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics*, 17(2):213–240.

Torgler, B. (2005). Tax morale in latin america. *Public Choice*, 122(1):133–157.