

Social Welfare Equilibrium*

Stan Cheung[†]

26th October, 2026

Click [here](#) for the latest draft

Abstract

Policies can respond to preferences, but preferences can also respond to policies. Endogenous preferences lead to both challenges and opportunities for policy making. A policy that induces preferences which tighten incentive constraints can severely contract the options available to a policy maker. But the converse can greatly expand these options. We propose the concept of a *social welfare equilibrium* for policy evaluation under endogenous preferences. We show that social welfare equilibria also offer a novel way to compare social welfare functions. We illustrate such a comparison in an application to optimal income taxation. We find that more inequality-averse social welfare functions induce more efficient social welfare equilibria, which suggests that equality and efficiency can be complements, rather than substitutes, when preferences are endogenous.

JEL: D63; D71; H21.

Keywords: Endogenous preferences, social welfare functions, social welfare equilibrium, optimal income taxation, efficiency-equality trade-off.

*I would like to thank my supervisors, Marco Mariotti and Roberto Veneziani, for their invaluable guidance and support, as well as Douglas Bernheim for his generous hospitality during a visit to Stanford in the Spring Quarter 2026. I am also grateful to Kristoffer Berg, Gabriel Carroll, Roberto Corrao, Piotr Dworczak, Andrea Gallo, Alberto Gazzi, Nicolas Gravel, Klajdi Hoxha, Asen Ivanov, Matthew Jackson, Jan Knoepfle, Fereshteh Mohseninejad, Ludovic Renou, Ariel Rubinstein, Evgenii Safonov, Enrico Mattia Salonia, Edward Schlee, Ran Spiegler, Ellen Stuart, Ester Sudano, Filip Tokarski, Zhang Xu, Gabriel Ziegler, and participants at PET 2025, AETW 2026, SSCW 2026, the 2026 Econometric Society Asian School on Economic Theory, the 2026 Berlin PhD Conference in Economics, and GRASS XX for helpful comments and discussions. Financial support from the ESRC (ES/P000703/1 2743394) is gratefully acknowledged.

[†]School of Economics and Finance, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom (stan.cheung@qmul.ac.uk).

1 Introduction

Policies can respond to preferences, but preferences can also respond to policies. These forms of endogenous preferences lead to both challenges and opportunities for policy making. A naive policy maker can choose policies that unintentionally induce preferences which tighten their incentive constraints, severely contracting the options available to them. But a sophisticated policy maker can choose policies that intentionally loosen incentive constraints, greatly expanding their available options.

Take, for example, the large-scale increase in fiscal capacity in most developed countries during the twentieth century. An increase in voluntary tax compliance, driven by policies that prioritised public good provision, played a pivotal role alongside improvements in enforcement to increase tax revenues from around 10% of GDP in most countries in the 1900s, to over 40% in some by the 2000s (e.g. Besley 2020). This increase in tax revenues has enabled governments to expand the scope of their policies into sectors that had been beyond reach, such as the provision of public education, health care, and pensions. Policies can also provoke retaliatory responses when they are deemed unfair (e.g. Besley, Jensen, et al. 2023), signal the intentions of policy makers that prompt individuals to reciprocate (e.g. Fehr, Hart, et al. 2011), generate social interactions that shape other-regarding preferences (e.g. Rao 2019), introduce legislation that shift social norms (e.g. Aksoy et al. 2020), and share information that change the beliefs used to form expected payoffs (e.g. Baseler et al. 2025), among many further examples.¹

This paper proposes the concept of a *social welfare equilibrium* for policy evaluation when preferences are endogenous. In the standard social welfare function framework, a policy maker evaluates policies based on the allocations they induce, taking individual preferences as given. Policy evaluation is based on the concept of an *optimal allocation*, which maximises social welfare over a given set of feasible allocations. However, when preferences can respond to the policies chosen by a policy maker, choosing an optimal allocation can change not only the way in which individuals value allocations, but also which allocations are feasible at all, through the incentive constraints. Choosing an initially optimal allocation can induce new preferences under which that allocation becomes suboptimal, perhaps not even feasible. But choosing an initially suboptimal allocation can induce preferences under which there are new optimal allocations that were not even feasible. Evaluating policies using the concept of optimal allocations can be misleading when preferences are endogenous.

A *social welfare equilibrium* is a utility profile and allocation that is possible for a policy maker to induce. At such an equilibrium, the allocation maximises a social welfare function given that utility profile, and the utility profile is stable given that allocation, or as we shall say, *does not respond* to that allocation. Specifically, a social welfare equilibrium, $(\hat{u}, \hat{x})$, of

¹See Bowles and Polanía-Reyes (2012) and Bowles (2016) for surveys.

a social welfare function W is a utility profile $\hat{u}$, together with an allocation $\hat{x}$, such that the policy maker is maximising W over all allocations that are feasible under $\hat{u}$ and to which $\hat{u}$ does not respond. The policy maker maximises social welfare only over allocations to which $\hat{u}$ does not respond, even when there are feasible allocations x under $\hat{u}$ that would increase social welfare. This is because the policy maker anticipates that choosing x would not be possible in the sense that it would lead to a response in $\hat{u}$. Specifically, individuals have an initial utility profile based on which the policy maker chooses an allocation. However, this choice of allocation can induce individuals to change their utility profile. Therefore, the new utility profile induced might not be the initial profile that the policy maker had in mind.

Social welfare equilibria may be used for policy evaluation by giving a policy maker all the utility profiles and allocations that are possible for them to induce. The policy maker may rank social welfare equilibria based on their normative criteria. They may use these rankings to evaluate policies based on the social welfare equilibria any given policy would induce. They anticipate that choosing optimal allocations can induce normatively undesirable social welfare equilibria, but choosing suboptimal allocations can induce social welfare equilibria that are normatively desirable.

Consider, for example, a setting with two individuals and one good. Figure 1(a) shows the feasible allocations $X(u)$ under the utility profile u . Choosing an initially optimal allocation $m_W(u)$ can induce the individuals to have a new utility profile $\hat{u}$ in Figure 1(b) where $m_W(u)$ cannot be optimal because it is not even feasible. But choosing an initially suboptimal allocation s can induce another utility profile $\hat{u}'$ in Figure 1(c) which has an optimal allocation $m_W(\hat{u}')$ that was not even feasible under u . Suppose that choosing $m_W(\hat{u})$ under $\hat{u}$, as well as $m_W(\hat{u}')$ under $\hat{u}'$, does not induce responses in preferences. Then $(\hat{u}, m_W(\hat{u}))$ and $(\hat{u}', m_W(\hat{u}'))$ are social welfare equilibria. A sophisticated policy maker facing a utility profile u may use the set of social welfare equilibria $\{(\hat{u}, m_W(\hat{u})), (\hat{u}', m_W(\hat{u}'))\}$ to evaluate their choice of allocation. They may rank the social welfare equilibria based on their normative desirability, anticipate that choosing allocation $m_W(u)$ would induce $(\hat{u}, m_W(\hat{u}))$ but choosing s would induce $(\hat{u}', m_W(\hat{u}'))$, and choose the allocation that induces the more desirable equilibrium. A naive policy maker who only uses the concept of an optimal allocation would always only choose $m_W(u)$ which would induce the equilibrium $(\hat{u}, m_W(\hat{u}))$ that may not always be as desirable as $(\hat{u}', m_W(\hat{u}'))$.

Social welfare equilibria also offer a novel way to compare social welfare functions. Specifically, a social welfare function may be compared with another based on the social welfare equilibria they induce, instead of only on their intrinsic properties. Such a comparison is illustrated in an application to optimal income taxation. We find that a social welfare function is at least as inequality-averse as another if and only if it induces at least as efficient social welfare equilibria over optimal income taxation models. This suggests that efficiency and equality, despite often being treated as substitutes in optimal

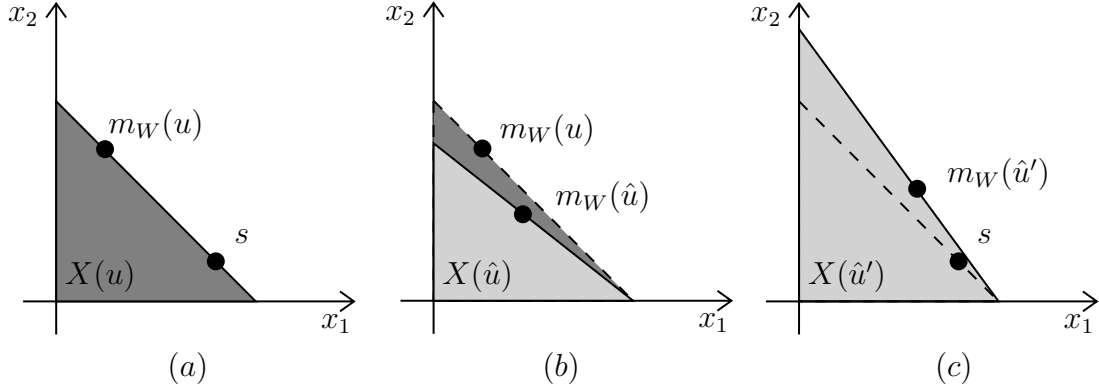


Figure 1: A setting with two individuals, $\{1, 2\}$, and one good, x , where a policy maker uses a social welfare function W . Figure (a) is the setting under utility profile u , Figure (b) under $\hat{u}$, and Figure (c) under $\hat{u}'$.

income taxation models, can be complements when preferences are endogenous.

1.1 Related literature

This paper generalises the standard Bergson (1938) and Samuelson (1938) social welfare function framework by incorporating preferences that can endogenously respond to the allocations chosen by a policy maker.² We propose an equilibrium concept that can be used for policy evaluation under the generalised framework. We also show that these equilibria offer a novel way to compare social welfare functions, leading to an alternative characterisation of the maximin social welfare function over optimal income taxation models that, unlike most characterisations (e.g. Hammond 1976; Bosmans and Ooghe 2013), does not depend on equity properties, but only on a single efficiency property. Our characterisation can be compared to Mariotti and Veneziani (2009) in that it uses such an efficiency property instead of a liberal non-interference property.

The generalisation of the social welfare function framework in this paper has a similar flavour to Hahn (1982) and Landsburg (2007) who suggest that individuals also have preferences over the social welfare function used by a policy maker. Individuals in our framework have preferences that can respond to the allocations chosen by a policy maker who maximises a given social welfare function, instead of preferences over the collection of social welfare functions themselves.

We also contribute to the literature on incorporating other-regarding preferences into welfare economics. A strand of this literature incorporates fixed social preferences into the general equilibrium model in a pure exchange setting (e.g. Kranich 1988; Dufwenberg et al. 2011; Bonnisseau et al. 2023). Another strand characterises social welfare functions that aggregate utility functions representing fixed other-regarding preferences (e.g. Decerf and

²See Weymark (2016) for a survey of the standard social welfare function framework.

Van der Linden 2016; Schmidt and Wichardt 2019; Treibich 2019; Berg and Piacquadio 2023). We incorporate other-regarding preferences that are not fixed, but can endogenously respond to the allocation chosen by a policy maker.

Our framework is comparable to those used in implementation theory (e.g. Jackson 2001), as well as models of endogenous preferences for specific applications such as public good contributions (Bar-Gill and Fershtman 2005; Bowles and Hwang 2008; Hwang and Bowles 2014), fiscal capacity (Besley 2020), and climate policies (Mattauch et al. 2022; Besley and Persson 2023). Our equilibrium concept pins down which utility functions individuals hold in their strategic interactions, as compared to the equilibrium concepts used for game forms which pin down which actions individuals choose in their strategic interactions. The equilibrium concepts used for game forms depend on our social welfare equilibria because the actions which individuals choose depend on the utility functions which individuals hold.

Our application is in the setting of optimal income taxation.³ We find that efficiency and equality, despite often being treated as substitutes in this literature, can be complements under endogenous preferences. Fleurbaey, Kanbur, et al. (2025) study efficiency and equality in terms of the outcomes of a general equilibrium model embedded in a social game, showing that incorporating these social interactions enriches the notions of efficiency and equality with implications for the ways of achieving them. We study efficiency and equality instead as properties on social welfare functions, showing that they are equivalent over optimal income taxation models when preferences are endogenous.

The paper is structured as follows. Section 2 presents the framework. Section 3 proposes the concept of a social welfare equilibrium, as well as conditions under which they exist. Sections 4 and 5 illustrate a comparison of social welfare functions based on the social welfare equilibria they induce in an application to optimal income taxation. Section 4 outlines the setting of an optimal income taxation model. Section 5 shows that more inequality-averse social welfare functions are characterised by more efficient social welfare equilibria over optimal income taxation models. Section 6 concludes.

2 Preliminaries

Let $\mathcal{I} \subseteq \mathbb{N}$ be a set of $N > 0$ *individuals*. Let $\mathcal{G} \subseteq \mathbb{N}$ be a set of $M > 0$ *goods*. Denote by $\mathbb{R}_+$ the set of non-negative real numbers. The set of all *allocations* is given by $\mathbb{R}_+^{MN}$. An allocation is a vector where the value of each component is the amount of one good allocated to one individual. Let the vector inequalities be given by $x \gg y$ if and only if $x_i > y_i$ for all $i \in \mathcal{I}$.

³See Fleurbaey and Maniquet (2018) and Kaplow (2024) for surveys.

We incorporate endogenous preferences by endowing each individual with a set of utility functions that they might possibly hold, as well as with a response function that captures the way in which their utility function responds to the allocations chosen by a policy maker. A *utility space* consists of all the possible profiles of utility functions the individuals may hold.

Definition 2.1. *A utility space is given by*

$$U = \prod_{i \in \mathcal{I}} U_i$$

that is compact, convex, and equipped with the product topology, where each U_i is a set of continuous utility functions, $u_i : \mathbb{R}_+^{MN} \rightarrow \mathbb{R}$, that is equipped with the topology of uniform convergence on compact sets.

Utility spaces comprise profiles of utility functions that may differ in the way they make trade-offs across different goods, as well as across different individuals. Some examples of preferences follow.

Example 2.1. (*Selfish Preferences*). *Let $u_i(x) = v_i(x_i)$ for some continuous $v_i : \mathbb{R}_+^M \rightarrow \mathbb{R}$.*

Example 2.2. (*Constant Elasticity of Substitution Preferences*). *Let*

$$u_i(x) = \begin{cases} \left(\sum_{j \in \mathcal{I}} \alpha_j v_j(x_j)^\rho \right)^{1/\rho} & \text{whenever } \rho \neq 0, \\ \prod_{j \in \mathcal{I}} v_j(x_j)^{\alpha_j} & \text{whenever } \rho = 0, \end{cases}$$

for some $\rho \in (-\infty, 1]$, $\alpha \in \mathbb{R}_+^N$, and continuous $v_j : \mathbb{R}_+^M \rightarrow \mathbb{R}$ (e.g. Edgeworth 1881; Andreoni and Miller 2002; Fisman et al. 2007).

Example 2.3. (*Fairness Preferences*). *Let*

$$u_i(x) = w_i(v_i(x_i), f_i(x))$$

for some continuous and increasing $v_i : \mathbb{R}_+^M \rightarrow \mathbb{R}$, continuous $f_i : \mathbb{R}_+^{MN} \rightarrow \mathbb{R}^{K_i}$ that captures deviations from the $K_i \leq N$ fairness benchmarks of i , and continuous $w_i : \mathbb{R} \times \mathbb{R}^{K_i} \rightarrow \mathbb{R}$ such that $w_i(v, \phi) \leq w_i(v, 0)$ for all $v \in \mathbb{R}$ and $\phi \in \mathbb{R}^{K_i}$ (e.g. Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Cappelen et al. 2007).

A *response function* captures the way in which the utility function of an individual responds to the allocations chosen by a policy maker.

Definition 2.2. A response function of an individual i is given by

$$R_i : U \times \mathbb{R}_+^{MN} \rightarrow U_i$$

that is continuous in both components.

A response function R_i maps an initial utility profile u and an allocation chosen by the policy maker x to a new utility function $R_i(u, x)$. This means that the initial utility function of an individual u_i may respond to the allocation chosen by the policy maker x directly, but also indirectly through the initial utility functions of other individuals $(u_j)_{j \neq i}$.

Response functions encompass a range of ways in which preferences might respond to allocations.

Example 2.4. (*Reciprocal Voluntary Tax Compliance*). Let

$$R_i(u, x)(y) = \theta'_i u_i^V(y) + (1 - \theta'_i) u_i^M(y)$$

where $u_i^V, u_i^M : \mathbb{R}_+^{MN} \rightarrow [0, 1]$ are civic-minded and materialistic utility functions whose weightings are governed by

$$\theta'_i = \begin{cases} \theta_i + (1 - \theta_i) \sigma(u) (u_i^V(x) - u_i^M(x)) & \text{if } u_i^V(x) - u_i^M(x) \geq 0 \\ \theta_i + \theta_i (1 - \sigma(u)) (u_i^V(x) - u_i^M(x)) & \text{if } u_i^V(x) - u_i^M(x) < 0 \end{cases}$$

where $\theta, \theta' \in [0, 1]$ and $\sigma(u) = \frac{1}{N} \sum_{j \in \mathcal{I}} \theta_j$ is the share of civic-mindedness (e.g. Besley 2020; see also Schulz et al. 2026).

Example 2.5. (*Social Learning*). Let

$$R_i(u, x)(y) = \int_{\Omega} v_i(y, \omega) d\mu'_i(\omega)$$

where Ω is the state space, $v_i : \mathbb{R}_+^{MN} \times \Omega \rightarrow \mathbb{R}$ a state-dependent utility function, and $\mu'_i : \Omega \rightarrow [0, 1]$ a probability distribution over Ω such that

$$\mu'_i = \sum_{j \in \mathcal{I}} p_{ij} \mu_j$$

where $p_{ij} \in [0, 1]$ are weights given by i to j such that $\sum_{j \in \mathcal{I}} p_{ij} = 1$ (e.g. DeGroot 1974; see also Mobius and Rosenblat 2014).

Example 2.6. (*Habit Formation*). Let

$$R_i(u, x)(y) = \sum_{g \in \mathcal{G}} \alpha_{ig} f_{ig}(y_{ig} - h_{ig}(x_{ig}))$$

where $\alpha_{ig} > 0$, $f_{ig} : \mathbb{R} \rightarrow \mathbb{R}$ is continuously differentiable, increasing, and strictly concave, and $h_{ig} : \mathbb{R}_+ \rightarrow [0, H_{ig}]$ is a continuous and non-decreasing habit-formation function for individual i and good g for some $H_{ig} < \infty$ (e.g. Pollak 1970; see also Frederick et al. 2002).

The ways in which utility functions respond to allocations affect which allocations are feasible. This is because the feasibility of implementing an allocation depends not only on resources and technologies, but also on incentive compatibility. Incentive constraints change when utility functions change. They can loosen under a utility profile, making some allocations feasible, but tighten under another utility profile, making them unfeasible.

Definition 2.3. A feasible allocation correspondence is given by

$$X : U \rightrightarrows \mathbb{R}_+^{MN}$$

that is continuous, compact-valued, and such that $\cap_{u \in U} X(u) \neq \emptyset$.

Feasible allocation correspondences encompass a range of ways in which the feasibility of an allocation might be constrained by the utility functions of individuals.

Example 2.7. (*Participation Constraints*). Let

$$X(u) = \{x \in A \mid u_i(x) \geq u_i(x_0) \text{ for all } i \in \mathcal{I}\}$$

where A is compact and x_{i0} is an outside option for all $i \in \mathcal{I}$.

Example 2.8. (*Envy-Freeness Constraints*). Let

$$X(u) = \{x \in A \mid u_i(x) \geq u_i(d_i(x)) \text{ for all } i \in \mathcal{I}\}$$

where A is compact and $d_i(x)$ gives the allocation that is realised when i deviates from x for all $i \in \mathcal{I}$.

Example 2.9. (*Coalitional-Stability Constraints*). Let

$$X(u) = \{x \in A \mid \text{for all } C \subseteq \mathcal{I} \text{ and } z \in D_C(x), \text{ there is } i \in C \text{ such that } u_i(x) \geq u_i(z)\}$$

where A is compact and C is a non-empty coalition and $D_C(x) \subseteq \mathbb{R}_+^{MN}$ gives the allocations that can be realised when coalition C deviates from x .

A social welfare function $W : \mathbb{R}^N \rightarrow \mathbb{R}$ gives the social welfare $W(v)$ for all welfare profiles $v = u(x)$ where $u(x)$ is a utility profile u under an allocation x . The function W is continuous if $W^{-1}(V)$ is open for all $V \subseteq \mathbb{R}$ open. It is Paretian if $W(v) > W(v')$ whenever $v \gg v'$. Let $\mathcal{W}$ be the collection of all continuous Paretian social welfare functions.

<i>Standard Framework Framework Under Endogenous Preferences</i>		
<i>Preferences</i>	Given by u	Given by u in a <i>utility space</i> U subject to a <i>response profile</i> R
<i>Feasible Allocations</i>	Given by a set X	Given by a <i>correspondence</i> X over utility profiles u
<i>Social Welfare Functions</i>	Given by W over welfare profiles $v = u(x)$	

Table 1: Comparison of the standard social welfare function framework and the framework under endogenous preferences.

Let $\mathcal{U}$ be the collection of all utility spaces, $\mathcal{R}$ the set of all response profiles, and $\mathcal{X}$ the collection of all feasible allocation correspondences. A *setting* is a tuple $S = (U, R, X) \in \mathcal{U} \times \mathcal{R} \times \mathcal{X}$. Denote by $\mathcal{S}$ the domain of all settings. The standard framework is a setting where the utility space is a singleton. Table 1 compares the standard framework and the framework under endogenous preferences.

3 Social Welfare Equilibria

Policy evaluation in the standard social welfare function framework is based on the concept of an *optimal allocation*. Policy makers take the profiles of utility functions u_0 as given in order to maximise a social welfare function W over a given set of feasible allocations X_0 ,

$$M_W := \arg \max_{x \in X_0} W(u_0(x)).$$

Policy makers take the preferences of individuals, as well as the feasible allocations, as given. Individuals take the chosen allocation as given.

Optimal allocations are a misleading concept for policy evaluation when preferences can endogenously respond to the allocations chosen by the policy maker. This is because the optimal allocations become a correspondence that depends on the utility profiles, $M_W : U \rightrightarrows \mathbb{R}_+^{MN}$, given by

$$M_W(u) := \arg \max_{x \in X(u)} W(u(x))$$

for all $u \in U$. Choosing an optimal allocation $m_W(u)$ under an initial utility profile u

can lead to a response of the utility functions, $R(u, m_W(u)) = u'$, that makes the chosen allocation suboptimal, $m_W(u) \notin M_W(u')$, perhaps even unfeasible, $m_W(u) \notin X(u')$.

We propose the concept of a *social welfare equilibrium* for policy evaluation under endogenous preferences.

Definition 3.1. A social welfare equilibrium $(\hat{u}, \hat{x})$ of a social welfare function W is a utility profile $\hat{u}$, together with an allocation $\hat{x}$, such that:

(i) The allocation $\hat{x}$ is feasible,

$$\hat{x} \in X(\hat{u}).$$

(ii) The utility profile $\hat{u}$ does not respond to $\hat{x}$,

$$R(\hat{u}, \hat{x}) = \hat{u}.$$

(iii) Social welfare is maximised among the allocations under $\hat{u}$ that are feasible and to which $\hat{u}$ does not respond,

$$W(\hat{u}(\hat{x})) \geq W(\hat{u}(x)) \text{ for all } x \text{ such that } x \in X(\hat{u}) \text{ and } R(\hat{u}, x) = \hat{u}.$$

A social welfare equilibrium is a utility profile and allocation that is possible for a policy maker in the sense that choosing that allocation under that utility profile would not lead preferences to respond. At a social welfare equilibrium $(\hat{u}, \hat{x})$, the allocation $\hat{x}$ maximises social welfare only over allocations to which $\hat{u}$ does not respond, even when there are feasible allocations x under $\hat{u}$ that have higher social welfare. This is because the policy maker anticipates that choosing x cannot be possible because it would lead to a response in $\hat{u}$. Specifically, individuals have an initial utility profile based on which the policy maker chooses an allocation. However, this choice of allocation can induce individuals to change their utility profile. Therefore, the new utility profile induced might not be the initial profile that the policy maker had in mind.

Social welfare equilibria may be used for policy evaluation when preferences are endogenous. A policy maker may rank the set of all social welfare equilibria based on their normative criteria. They may use these rankings to evaluate policies based on the social welfare equilibria a given policy would induce. A policy maker may choose allocations that are not in social welfare equilibrium when they anticipate this would lead to responses in utility profiles that induce social welfare equilibria they find normatively desirable. There are many normative criteria a policy maker may use to rank social welfare equilibria.

Definition 3.2. A normative criterion over social welfare equilibria is a binary relation $\succsim$ over $U \times \mathbb{R}_+^{MN}$ such that $(\hat{u}, \hat{x}) \succsim (\hat{u}', \hat{x}')$ if and only if $(\hat{u}, \hat{x})$ is as normatively desirable as $(\hat{u}', \hat{x}')$.

A normative criterion that requires the utility functions of an individual to be comparable across utility profiles is to evaluate social welfare equilibria based only on their induced social welfare.

Example 3.1. (*Social-Welfare-Based Comparisons*). Let $\succsim$ be such that

$$(\hat{u}, \hat{x}) \succsim (\hat{u}', \hat{x}') \text{ if and only if } W(\hat{u}(\hat{x})) \geq W(\hat{u}'(\hat{x}')).$$

There are also normative criteria that do not require the comparability of utility across different profiles.

Example 3.2. (*Dominance-Based Comparisons*). Let $\succsim$ be such that

$$(\hat{u}, \hat{x}) \succsim (\hat{u}', \hat{x}') \text{ if and only if } W(\hat{u}(\hat{x})) \geq W(\hat{u}(\hat{x}')) \text{ and } W(\hat{u}'(\hat{x})) \geq W(\hat{u}'(\hat{x}'))$$

(e.g. Schoeffler 1952; Weisbrod 1977).

Example 3.3. (*Feasibility-Based Comparisons*). Let $\succsim$ be such that

$$(\hat{u}, \hat{x}) \succsim (\hat{u}', \hat{x}') \text{ if and only if } X(\hat{u}) \subseteq X(\hat{u}')$$

(e.g. Sugden 2018).

These are meaningful rankings because social welfare equilibria exist for all continuous Paretian social welfare functions.

Theorem 1. (Existence). For all $S \in \mathcal{S}$ and $W \in \mathcal{W}$, the set of all social welfare equilibria Σ_W of W is non-empty.

Proof. Take any $S \in \mathcal{S}$ and $W \in \mathcal{W}$. First, we find a utility profile $\hat{u}$ such that $R(\hat{u}, x) = \hat{u}$ for some $x \in X(\hat{u})$. Because $\cap_{u \in U} X(u) \neq \emptyset$, there is some $x^* \in \mathbb{R}_+^{MN}$ such that $x^* \in X(u)$ for all $u \in U$. Therefore, $s : U \rightarrow \mathbb{R}_+^{MN}$ such that $s(u) = x^*$ for all $u \in U$ is a continuous selector of X . Because R is continuous, we have that $\gamma : U \rightarrow U$ such that $\gamma(u) := R(u, s(u))$ for all $u \in U$ is continuous. Because U is compact and convex, by the Schauder fixed-point theorem, there is $\hat{u} \in U$ such that $\gamma(\hat{u}) = \hat{u}$.

Second, we find an allocation $\hat{x} \in \mathbb{R}_+^{MN}$ such that $(\hat{u}, \hat{x})$ is a social welfare equilibrium of W . Let $F : U \rightrightarrows \mathbb{R}_+^{MN}$ be a correspondence given by $F(u) = \{x \in \mathbb{R}_+^{MN} \mid R(u, x) = u\}$ for all $u \in U$. We show that F is closed-valued. Take any $u \in U$ and $\{x_n\}_{n \in \mathbb{N}} \subset F(u)$ such that $x_n \rightarrow x$ as $n \rightarrow \infty$. Because R is continuous, and because $R(u, x_n) = u$ for all $n \in \mathbb{N}$, we have $R(u, x) = u$ so that $x \in F(u)$. Therefore, because the intersection of closed sets is closed, X is compact-valued, and $F(u)$ is closed-valued, we have that $X(u) \cap F(u)$ is compact for all $u \in U$. Moreover, $X(\hat{u}) \cap F(\hat{u})$ is non-empty because $x^* \in X(\hat{u}) \cap F(\hat{u})$. Therefore,

because $\hat{u}$ and W are continuous, by the extreme value theorem, $\arg \max_{x \in X(\hat{u}) \cap F(\hat{u})} W(\hat{u}(x))$ is non-empty. Take any $\hat{x} \in \arg \max_{x \in X(\hat{u}) \cap F(\hat{u})} W(\hat{u}(x))$. We have that $(\hat{u}, \hat{x})$ is a social welfare equilibrium so that Σ_W is non-empty. $\square$

Social welfare equilibria also offer a novel way to compare social welfare functions. Specifically, it becomes possible to compare a social welfare function W with another W' based on the social welfare equilibria Σ_W and $\Sigma_{W'}$ they induce instead of only based on intrinsic properties of W and W' . We illustrate these comparisons in an application to optimal income taxation.

4 Optimal Income Taxation

An *optimal income taxation model* is a setting where there are only two goods, *consumption* $x_i^1 \in [0, \infty)$, and *leisure* $x_i^2 \in [0, 1]$. Denote $A = ([0, \infty) \times [0, 1])^N$. Consumption is public information, but leisure is private information. Each individual i is endowed with a *productivity*, $w_i \geq 0$, such that $w_i < w_{i+1}$ for all $i < N$. Productivity is private information, but the distribution of productivity across individuals is public information.

Individuals are endowed with a set of utility functions U_i such that the utility space U is *regular* in that for all $\gamma \in \mathbb{R}$ and $x, y \in \mathbb{R}_+^{2N}$ such that $x \gg y$, we have $u_i(x) \geq \gamma$ for all $i \in \mathcal{I}$ whenever $u_i(y) \geq \gamma$ for all $i \in \mathcal{I}$. Regular utility spaces ensure that whenever an allocation y improves to x such that $x \gg y$, this does not lead to a decrease in utility for at least the individual who had the lowest utility under x . This individual is in a sense self-regarding enough to weakly prefer having an improved bundle regardless of the improvements to others. Utility spaces where all individuals have homogeneous purely self-regarding utility functions are regular. Individuals are also endowed with response functions $R_i : U \times \mathbb{R}_+^{2N} \rightarrow U_i$ for all $i \in \mathcal{I}$ where U is regular.

Feasible allocations are given by

$$X(u) = X_1 \cap X_2(u)$$

for all $u \in U$ where $X_1 \subset A$ is a set of *first-best feasible allocations* and $X_2 : U \rightrightarrows A$ is a *second-best feasible allocation correspondence*. First-best feasible allocations are given by the budget constraint of the government. A first-best feasible allocation is not implementable when there is an individual who has an incentive to imitate another individual under that allocation. This is possible because productivity is private information. The second-best feasible allocations are given by the allocations in which no individual has the incentive to imitate any other individual. Formally, the *income* of an individual i is given by $y_i := w_i(1 - x_i^2)$. The government has a budget constraint such that the sum of its taxes cannot

be less than some $G \in \mathbb{R}$. This gives the *first-best feasible allocations*,

$$X_1 := \left\{ x \in A \mid \sum_{i \in \mathcal{I}} y_i - x_i^1 \geq G \right\}.$$

A first-best feasible allocation is implementable when no individual has an incentive to imitate another individual under that allocation. This gives the *second-best feasible allocations*,

$$X_2(u) := \{x \in A \mid u_i(x) \geq u_i(d_{ij}(x)) \text{ for all } i, j \in \mathcal{I} \text{ such that } w_i \geq y_j\},$$

for all $u \in U$ where $d_{ij}(x)$ gives the allocation that is realised when i imitates j under x ,

$$(d_{ij}(x))_i = \left(x_j^1, 1 - \frac{y_j}{w_i} \right) \text{ and } (d_{ij}(x))_k = x_k \text{ for all } k \neq i.$$

An optimal income taxation model is summarised as follows.

Definition 4.1. An optimal income taxation model is a setting (U, R, X) where there are two goods consumption $x_i^1 \in [0, \infty)$ and leisure $x_i^2 \in [0, 1]$, the utility space U is regular, the response profile is given by $R : U \times \mathbb{R}_+^{2N} \rightarrow U$, and the feasible allocations are given by $X(u) = X_1 \cap X_2(u)$ for all $u \in U$ for some first-best feasible allocation set X_1 and second-best feasible allocation correspondence X_2 .

Denote by $\mathcal{T} \subset \mathcal{S}$ the domain of all optimal income taxation models. Section 5 shows that efficiency and equality can be complements over $\mathcal{T}$, not substitutes.

5 Efficiency and Equality

Efficiency and equality are often treated as substitutes, not complements, in optimal income taxation. We find that they can be complements when preferences are endogenous. Specifically, we show that ranking social welfare functions based on the efficiency of the social welfare equilibria they induce is equivalent to ranking them based on their degree of inequality aversion.

Definition 5.1. A social welfare function W is at least as efficient as another W' over $\mathcal{T}$ when for all $T \in \mathcal{T}$ and $(\hat{u}', \hat{x}') \in \Sigma_{W'}$, there is $(\hat{u}, \hat{x}) \in \Sigma_W$ such that $\hat{x}' \not\gg \hat{x}$.

A social welfare function W is at least as efficient as W' over optimal income taxation models when for all such settings and social welfare equilibria $(\hat{u}', \hat{x}')$ induced by W' , there is a social welfare equilibrium of W whose allocation $\hat{x}$ is not strictly dominated by the allocation $\hat{x}'$ in the sense that $\hat{x}' \gg \hat{x}$.

This is a relatively weak form of efficiency. Indeed, strict domination of $\hat{x}$ by $\hat{x}'$ means that every good to every individual in $\hat{x}'$ is strictly greater than that of $\hat{x}$. For W to be at least as efficient as W' , all social welfare equilibria of W' need only have one corresponding social welfare equilibrium of W where at least one individual is allocated at least as much of at least one good. Conversely, W' not at least as efficient as W would mean that there is an optimal income taxation model and a social welfare equilibrium of W whose allocation gives strictly more of every good to every individual than the allocations of all social welfare equilibria of W' .

Despite seeming relatively weak, we find that this ranking is equivalent to a seemingly much stronger ranking of social welfare functions based on their degree of inequality aversion. Let $\underline{W}$ be the maximin social welfare function given by $\underline{W}(v) := \min_{i \in \mathcal{I}} v_i$ for all $v \in \mathbb{R}^N$. This social welfare function represents the case of an infinite degree of inequality aversion.

Definition 5.2. *A social welfare function W is at least as inequality averse as another W' when for all $a, b \in \mathbb{R}^N$ such that $\underline{W}(a) \geq \underline{W}(b)$, we have $W(a) \geq W(b)$ whenever $W'(a) \geq W'(b)$.*

A social welfare function W is at least as inequality averse as W' when it is closer to $\underline{W}$ in the sense that whenever $\underline{W}$ and W' agree on the comparisons of utility profiles, W must also agree with $\underline{W}$. Ranking social welfare functions based on the efficiency of the social welfare equilibria they induce is equivalent to ranking them based on their degree of inequality aversion.

Theorem 2. (Complementarity of Efficiency and Equality for Optimal Income Taxation). *For all $W, W' \in \mathcal{W}$, we have that W is at least as efficient as W' over $\mathcal{T}$ if and only if W is at least as inequality averse as W' .*

Proof. For the ‘if’ part, take any $W, W' \in \mathcal{W}$ such that W is at least as inequality averse as W' . Suppose by contradiction that W is not at least as efficient as W' over $\mathcal{T}$. Therefore, there is $T \in \mathcal{T}$ and $(\hat{u}', \hat{x}') \in \Sigma_{W'}$ such that

$$\hat{x}' \gg \hat{x} \text{ for all } (\hat{u}, \hat{x}) \in \Sigma_W. \quad (1)$$

Because $(\hat{u}', \hat{x}') \in \Sigma_{W'}$, we have $R(\hat{u}', \hat{x}') = \hat{u}'$. Using the same argument as in the proof of Theorem 1, $C := X(\hat{u}') \cap \{x \in \mathbb{R}_+^{2N} \mid R(\hat{u}', x) = \hat{u}'\}$ is compact. Moreover, because $(\hat{u}', \hat{x}') \in \Sigma_{W'}$, we have $\hat{x}' \in C$ so that C is non-empty. Because $\hat{u}'$ and W are continuous, the composition $W \circ \hat{u}'$ is continuous in x . Therefore, by the extreme value theorem, $\arg \max_{x \in C} W(\hat{u}'(x))$ is non-empty.

Take any $(\hat{u}', x^*)$ such that $x^* \in \arg \max_{x \in C} W(\hat{u}'(x))$. Therefore, $(\hat{u}', x^*) \in \Sigma_W$ so that $\hat{x}' \gg x^*$ by (1). Because U is regular and $\hat{x}' \gg x^*$, we have $\min_{i \in \mathcal{I}} \hat{u}'_i(\hat{x}') \geq \min_{i \in \mathcal{I}} \hat{u}'_i(x^*)$ so that $\underline{W}(\hat{u}'(\hat{x}')) \geq \underline{W}(\hat{u}'(x^*))$. Moreover, because $(\hat{u}', \hat{x}') \in \Sigma_{W'}$ and $x^* \in C$, we have

$W'(\hat{u}'(\hat{x}')) \geq W'(\hat{u}'(x^*))$. Therefore, because W is at least as inequality averse as W' , we have $W(\hat{u}'(\hat{x}')) \geq W(\hat{u}'(x^*))$ so that $(\hat{u}', \hat{x}') \in \Sigma_W$. By (1), $\hat{x}' \gg \hat{x}$, a contradiction.

For the ‘only if’ part, take any $W, W' \in \mathcal{W}$ such that W is not at least as inequality averse as W' . We construct $T \in \mathcal{T}$ and $(\hat{u}', \hat{x}') \in \Sigma_{W'}$ such that $\hat{x}' \gg \hat{x}$ for all $(\hat{u}, \hat{x}) \in \Sigma_W$. Suppose the productivities of the individuals are given by w_i such that $w_i < w_{i+1}$ for all $i < N$. Because W is not at least as inequality averse as W' , there are $a, b \in \mathbb{R}^N$ such that $\underline{W}(a) \geq \underline{W}(b)$ and $W(a) < W(b)$, but $W'(a) \geq W'(b)$.

First, we construct the utility space. Let $\underline{c}, c, \bar{c}, \underline{y}, \bar{y} \in \mathbb{R}$ be such that

$$0 < \underline{c} < \bar{c} < \underline{y} < \bar{y} < \min_{i \in \mathcal{I}} w_i.$$

For all $i \in \mathcal{I}$, let

$$\underline{x}_i = \left(\underline{c}, 1 - \frac{\bar{y}}{w_i} \right) \text{ and } \bar{x}_i = \left(\bar{c}, 1 - \frac{\underline{y}}{w_i} \right).$$

Moreover, let $x^* \in \mathbb{R}_+^{2N}$ be such that $\underline{x} \ll x^* \ll \bar{x}$ and $V_{\underline{x}} := \{x \in \mathbb{R}_+^{2N} \mid \|x - \underline{x}\| \leq r^*\}$ for some $r^* > 0$ such that $x \ll x^*$ for all $x \in V_{\underline{x}}$. For all $x \in \mathbb{R}_+^{2N}$, let

$$u_i^*(x) = \alpha(x)(b_i - \min_{i \in \mathcal{I}} b_i) + \beta(x)(a_i - \min_{i \in \mathcal{I}} b_i) + \min_{i \in \mathcal{I}} b_i$$

where

$$\begin{aligned} \alpha(x) &= \max \left\{ 0, 1 - \frac{\|x - \underline{x}\|}{r^*} \right\} \\ \beta(x) &= \min_{1 \leq k \leq 2N} \left\{ \min \left\{ 1, \max \left\{ 0, \frac{x_k - x_k^*}{\bar{x}_k - x_k^*} \right\} \right\} \right\}. \end{aligned}$$

By construction, u^* is continuous, $u^*(\underline{x}) = b$, $u^*(\bar{x}) = a$, and

$$u^*(x) \leq \begin{cases} b & \text{for all } x \in V_{\underline{x}} \\ a & \text{for all } x \notin V_{\underline{x}}. \end{cases} \quad (2)$$

Let $U := \{u^*\}$. By construction, U is regular because $\min_{i \in \mathcal{I}} u_i^*(x) = \beta(x)(\min_{i \in \mathcal{I}} a_i - \min_{i \in \mathcal{I}} b_i) + \min_{i \in \mathcal{I}} b_i$ is non-decreasing in x .

Second, we construct a constant response profile given by $R(u, x) = u^*$ for all $u \in U$ and $x \in \mathbb{R}_+^{2N}$.

Third, we construct the feasible allocation correspondence. For all $i, j \leq N$, let

$$(d_{ij}(x))_i = \left(x_j^1, 1 - \frac{y_j}{w_i} \right) \text{ and } (d_{ij}(x))_k = x_k \text{ for all } k \neq i.$$

Moreover, let $X_1 = \{x \in A \mid \sum_{i \in \mathcal{I}} y_i - x_i^1 \geq 0\}$ and $X_2(u) = \{x \in A \mid u_i(x) \geq u_i(d_{ij}(x)) \text{ for all } i, j \in \mathcal{I} \text{ such that } w_i \geq y_j\}$ for all $u \in U$. Let $X(u) = X_1 \cap X_2(u)$ for all

$u \in U$. By construction, $\underline{x}, \bar{x} \in X(u^*)$ because i imitating j makes the $\underline{x}$ and $\bar{x}$ unchanged for all $i, j \in \mathcal{I}$.

By construction $(U, R, X) \in \mathcal{T}$. Because W is Paretian, and by (2), we have,

$$\begin{aligned} W(u^*(x)) &\leq W(b) \text{ for all } x \in V_{\underline{x}} \text{ and} \\ W(u^*(x)) &\leq W(a) < W(b) \text{ for all } x \notin V_{\underline{x}} \end{aligned}$$

so that $(\hat{u}, \hat{x}) \in \Sigma_W$ only if $\hat{x} \in V_{\underline{x}}$. Similarly, because W' is Paretian, and by (2),

$$W'(u^*(x)) \leq W'(a) \text{ for all } x \in X(u^*).$$

Therefore, because $\bar{x} \in X(u^*)$ and $R(u^*, \bar{x}) = u^*$, we have $(u^*, \bar{x}) \in \Sigma_{W'}$. Because $\bar{x} \gg x$ for all $x \in V_{\underline{x}}$, we have that W is not at least as efficient as W' . $\square$

Theorem 2 suggests that efficiency and equality can be complements when preferences are endogenous. Ranking social welfare functions based on a seemingly weak form of efficiency turns out to be equivalent to ranking them based on a seemingly much stronger notion of inequality aversion. Theorem 2 also offers a characterisation of the maximin social welfare function over optimal income taxation models which does not use any equality properties but only a single efficiency property.

Definition 5.3. *A social welfare function W is efficient in $\mathcal{W}$ over $\mathcal{T}$ when it is at least as efficient as all $W' \in \mathcal{W}$ over $\mathcal{T}$.*

Corollary 1. (Characterisation of Maximin for Optimal Income Taxation). *A social welfare function W is efficient in $\mathcal{W}$ over $\mathcal{T}$ if and only if it represents the maximin ordering.*

Proof. For the ‘if’ part, $\underline{W}$ is at least as inequality averse as all $W \in \mathcal{W}$. By Theorem 2, this means that $\underline{W}$ is at least as efficient as all $W \in \mathcal{W}$ over $\mathcal{T}$. For the ‘only if’ part, take any $W \in \mathcal{W}$. Suppose W is at least as efficient as all $W' \in \mathcal{W}$ over $\mathcal{T}$. By Theorem 2, this means that W is at least as inequality averse as all $W' \in \mathcal{W}$. Therefore, for all $a, b \in \mathbb{R}^N$, $W(a) \geq W(b)$ if and only if $\underline{W}(a) \geq \underline{W}(b)$. $\square$

Theorem 2 and Corollary 1 suggest that efficiency and equality, despite often being treated as substitutes, can in fact be complements over optimal income taxation models when preferences are endogenous.

6 Conclusion

The possibilities for policy making widen significantly when preferences are endogenous. A naive policy maker can unintentionally induce preferences that tighten their incentive

constraints, contracting the options available to them. But a sophisticated policy maker can intentionally loosen these constraints, expanding their available options. This paper proposes the concept of a social welfare equilibrium for policy evaluation under endogenous preferences. We illustrate that social welfare equilibria also offer a novel way to compare social welfare functions in an application to optimal income taxation. We find that efficiency and equality, despite often being treated as substitutes in the optimal income taxation literature, can be complements when preferences are endogenous.

There are several directions for further work. From an empirical perspective, the implementation of the framework under endogenous preferences is possible under the same techniques as used for the standard framework. Specifically, to identify social welfare equilibria, the additional primitives that an analyst would be required to identify would only be a correspondence of *socially acceptable allocations*, $A : U \rightrightarrows \mathbb{R}_+^{MN}$, such that for all $u \in U$, we have that $A(u)$ is made up of all $x \in X(u)$ such that $R(u, x) = u$ (see also Richter and Rubinstein 2020). Utility functions may be estimated by an analyst using existing techniques. Moreover, to identify social welfare equilibria, the analyst would not be required to fully identify the response functions. A social welfare equilibrium may equivalently be expressed as $(\hat{u}, \hat{x})$ such that $\hat{x} \in A(\hat{u})$ where there is no $\hat{y} \in A(\hat{u})$ such that $W(\hat{u}(\hat{y})) > W(\hat{u}(\hat{x}))$, meaning that identifying the correspondence of socially acceptable allocations would be enough.

From a theoretical perspective, there are many further settings to apply our framework, on top of optimal income taxation models. General equilibrium models, for example, are recovered when individuals are endowed with initial endowments, ownership shares in firms endowed with production plans, and take market prices as given. This setting suggests a way to study the non-cooperative implementation of social welfare equilibria. Specifically, whether or not a social welfare equilibrium can be supported by some price vector in a decentralised competitive equilibrium (see also Schlee and Khan 2022). Moreover, our framework offers a precise way to analyse different approaches to welfare analysis when preferences are endogenous (e.g. Gintis 1972; Aaron 1994; Bowles 2016; Ashraf and Bandiera 2017). Different approaches to welfare analysis under endogenous preferences imply different normative criteria on social welfare equilibria. It is possible to analyse these different approaches through their different normative evaluations of social welfare equilibria.

Finally, the idea that there are interactions between policy makers and individuals during the policy making process can be extended far beyond the welfarist framework used in this paper. Preferences may respond not only to the allocations chosen by a policy maker, but also, among many other things, to procedures and processes followed by a policy maker, rights and liberties guaranteed by a policy maker, and public reasoning between policy makers and individuals. A deliberative process where both the collective choices of the government and the individual choices of citizens are subject to reasoned scrutiny has been

held to be a necessary condition for democratising the economy (e.g. Habermas 1996; Dryzek 2002; Sen 2009). Developing frameworks to facilitate these forms of reasoned scrutiny would require us to move far beyond only utility information.

References

- Aaron, H. J. (1994). ‘Distinguished lecture on economics in government: Public policy, values, and consciousness’. *Journal of Economic Perspectives* 8.2, pp. 3–21.
- Aksoy, C. G., C. S. Carpenter, R. De Haas, and K. D. Tran (2020). ‘Do laws shape attitudes? Evidence from same-sex relationship recognition policies in Europe’. *European Economic Review* 124, p. 103399.
- Andreoni, J. and J. Miller (2002). ‘Giving according to GARP: An experimental test of the consistency of preferences for altruism’. *Econometrica* 70.2, pp. 737–753.
- Ashraf, N. and O. Bandiera (2017). ‘Altruistic capital’. *American Economic Review* 107.5, pp. 70–75.
- Bar-Gill, O. and C. Fershtman (2005). ‘Public policy with endogenous preferences’. *Journal of Public Economic Theory* 7.5, pp. 841–857.
- Baseler, T., T. Ginn, R. Hakiza, H. Ogude-Chambert, and O. Woldemikael (2025). ‘Can redistribution change policy views?: Aid and attitudes toward refugees in Uganda’. *Journal of Political Economy* 133.9, pp. 2874–2917.
- Berg, K. and P. G. Piacquadio (2023). ‘Fairness and Paretian Social Welfare Functions’. *Working Paper*.
- Bergson, A. (1938). ‘A reformulation of certain aspects of welfare economics’. *The Quarterly Journal of Economics* 52.2, pp. 310–334.
- Besley, T. (2020). ‘State capacity, reciprocity, and the social contract’. *Econometrica* 88.4, pp. 1307–1335.
- Besley, T., A. Jensen, and T. Persson (2023). ‘Norms, enforcement, and tax evasion’. *Review of Economics and Statistics* 105.4, pp. 998–1007.
- Besley, T. and T. Persson (2023). ‘The political economics of green transitions’. *Quarterly Journal of Economics* 138.3, pp. 1863–1906.
- Bolton, G. E. and A. Ockenfels (2000). ‘ERC: A theory of equity, reciprocity, and competition’. *American Economic Review* 90.1, pp. 166–193.
- Bonnisseau, J.-M., E. L. del Mercato, and P. Siconolfi (2023). ‘Existence of an equilibrium in arrowian markets for consumption externalities’. *Journal of Economic Theory* 209, p. 105638.

- Bosmans, K. and E. Ooghe (2013). ‘A characterization of maximin’. *Economic Theory Bulletin* 1, pp. 151–156.
- Bowles, S. (2016). *The moral economy: why good incentives are no substitute for good citizens*. Yale University Press, New Haven.
- Bowles, S. and S.-H. Hwang (2008). ‘Social preferences and public economics: Mechanism design when social preferences depend on incentives’. *Journal of Public Economics* 92.8-9, pp. 1811–1820.
- Bowles, S. and S. Polanía-Reyes (2012). ‘Economic incentives and social preferences: substitutes or complements?’ *Journal of Economic Literature* 50.2, pp. 368–425.
- Cappelen, A. W., A. D. Hole, E. O. Sorensen, and B. Tungodden (2007). ‘The pluralism of fairness ideals: An experimental approach’. *American Economic Review* 97.3, pp. 818–827.
- Decerf, B. and M. Van der Linden (2016). ‘Fair social orderings with other-regarding preferences’. *Social Choice and Welfare* 46, pp. 655–694.
- DeGroot, M. H. (1974). ‘Reaching a consensus’. *Journal of the American Statistical Association* 69.345, pp. 118–121.
- Dryzek, J. S. (2002). *Deliberative democracy and beyond: Liberals, critics, contestations*. Oxford University Press, Oxford.
- Dufwenberg, M., P. Heidhues, G. Kirchsteiger, F. Riedel, and J. Sobel (2011). ‘Other-regarding preferences in general equilibrium’. *The Review of Economic Studies* 78.2, pp. 613–639.
- Edgeworth, F. Y. (1881). *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. Kegan Paul, London.
- Fehr, E., O. Hart, and C. Zehnder (2011). ‘Contracts as reference points – experimental evidence’. *American Economic Review* 101.2, pp. 493–525.
- Fehr, E. and K. M. Schmidt (1999). ‘A theory of fairness, competition, and cooperation’. *Quarterly Journal of Economics* 114.3, pp. 817–868.
- Fisman, R., S. Kariv, and D. Markovits (2007). ‘Individual preferences for giving’. *American Economic Review* 97.5, pp. 1858–1876.
- Fleurbaey, M., R. Kanbur, and D. Snower (2025). ‘Efficiency and equity in a socially-embedded economy’. *Economic Theory* 79.1, pp. 1–56.

- Fleurbaey, M. and F. Maniquet (2018). ‘Optimal income taxation theory and principles of fairness’. *Journal of Economic Literature* 56.3, pp. 1029–1079.
- Frederick, S., G. Loewenstein, and T. O’Donoghue (2002). ‘Time discounting and time preference: A critical review’. *Journal of Economic Literature* 40.2, pp. 351–401.
- Gintis, H. (1972). ‘A radical analysis of welfare economics and individual development’. *The Quarterly Journal of Economics* 86.4, pp. 572–599.
- Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy*. MIT Press, Cambridge US.
- Hahn, F. (1982). ‘On some difficulties of the utilitarian economist’. In: *Utilitarianism and Beyond*. Ed. by A. Sen and B. Williams. Cambridge University Press.
- Hammond, P. J. (1976). ‘Equity, Arrow’s conditions, and Rawls’ difference principle’. *Econometrica: Journal of the Econometric Society*, pp. 793–804.
- Hwang, S.-H. and S. Bowles (2014). ‘Optimal incentives with state-dependent preferences’. *Journal of Public Economic Theory* 16.5, pp. 681–705.
- Jackson, M. O. (2001). ‘A crash course in implementation theory’. *Social Choice and Welfare* 18.4, pp. 655–708.
- Kaplow, L. (2024). ‘Optimal income taxation’. *Journal of Economic Literature* 62.2, pp. 637–738.
- Kranich, L. J. (1988). ‘Altruism and efficiency: A welfare analysis of the Walrasian mechanism with transfers’. *Journal of Public Economics* 36.3, pp. 369–386.
- Landsburg, S. E. (2007). ‘The methodology of normative economics’. *Journal of Public Economic Theory* 9.5, pp. 757–769.
- Mariotti, M. and R. Veneziani (2009). ‘“Non-interference” implies equality’. *Social Choice and Welfare* 32.1, pp. 123–128.
- Mattauch, L., C. Hepburn, F. Spuler, and N. Stern (2022). ‘The economics of climate change with endogenous preferences’. *Resource and Energy Economics* 69, no. 101312.
- Mobius, M. and T. Rosenblat (2014). ‘Social learning in economics’. *Annu. Rev. Econ.* 6.1, pp. 827–847.
- Pollak, R. A. (1970). ‘Habit formation and dynamic demand functions’. *Journal of Political Economy* 78.4, Part 1, pp. 745–763.
- Rao, G. (2019). ‘Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools’. *American Economic Review* 109.3, pp. 774–809.

- Richter, M. and A. Rubinstein (2020). ‘The permissible and the forbidden’. *Journal of Economic Theory* 188, no. 105042.
- Samuelson, P. A. (1938). ‘A note on the pure theory of consumer’s behaviour’. *Economica* 5.17, pp. 61–71.
- Schlee, E. E. and M. A. Khan (2022). ‘Money metrics in applied welfare analysis: a saddlepoint rehabilitation’. *International Economic Review* 63.1, pp. 189–210.
- Schmidt, U. and P. C. Wichardt (2019). ‘Inequity aversion, welfare measurement and the Gini index’. *Social Choice and Welfare* 52, pp. 585–588.
- Schoeffler, S. (1952). ‘Note on modern welfare economics’. *The American Economic Review* 42.5, pp. 880–887.
- Schulz, J. F., N. Nunn, and J. Henrich (2026). ‘Cultural Evolution in Economics’. *Working Paper*.
- Sen, A. (2009). *The Idea of Justice*. Belknap Press, Cambridge US.
- Sugden, R. (2018). *The community of advantage: A behavioural economist’s defence of the market*. Oxford University Press.
- Treibich, R. (2019). ‘Welfare egalitarianism with other-regarding preferences’. *Social Choice and Welfare* 52, pp. 1–28.
- Weisbrod, B. A. (1977). ‘Comparing utility functions in efficiency terms or, what kind of utility functions do we want?’ *The American Economic Review* 67.5, pp. 991–995.
- Weymark, J. A. (2016). ‘Social Welfare Functions’. In: *The Oxford Handbook of Well-Being and Public Policy, Volume 1*. Oxford University Press, New York.