1 Introduction

Recent developments in behavioural economics (BE) have deeply influenced the way governments design public policies. They give citizens access to online simulators to cope with tax and benefits systems and increasingly rely on *nudges* to guide individual decisions. Far beyond traditional instruments, the Covid-19 pandemic illustrates the large array of public health interventions available to support the vaccination campaign, from nudges to government mandates, including education, financial incentives and vaccination certificates with QR codes.[1] Finding the best policy requires an investigation of individual behaviours in order to understand the reasons for compliance.

The last decade has seen a surge in behavioural public finance (BPF). Beyond a taste for the foundation of public finance (PF) on cognitive sciences, this field of economic research is grounded on the conviction that a better understanding of individual behaviours could improve predictions of tax revenue and help design better-suited incentives to save for retirement, search for a new job, go to school or seek medical attention. Because behavioural agents may react in unpredictable ways, an examination of individual psychology quantitatively matters for the design of fiscal policies. Loss-averse US taxpayers facing a positive balance due on tax day are more eager to engage in tax-reducing activities, which results in a $1.4 billion loss of tax revenue (Rees-Jones, 2017). In contrast, procrastination on fiscal deductions costs US taxpayers around $50 million (Slemrod et al., 1997). A simple mailing sent by Bhargava and Manoli (2015) to inform 35,000 Californian tax filers about their eligibility to the Earned Income Tax Credit (EITC) led to refund claims totalling $4 million. The adoption of Electronic Toll Collection (ETC) raised tolls by 20 to 40 per cent because they are less salient for drivers (Finkelstein, 2009). The choice of a financially dominated health plan, which is more expensive regardless of how much care the employee requires, can cost each worker as much as $372 a year without additional health coverage (Bhargava et al., 2017).

Through a presentation of the most recent developments in BPF, this Element discusses the way BE has improved our understanding of fiscal policies. In contrast to traditional economic agents, behavioural agents have non-standard preferences or misperceive their economic environment. As a consequence, they may take different actions when faced with the same choice set. However, they can still be deemed rational since, conditional on modelling assumptions regarding their biases, their choices are consistent and predictable.

[1] On this topic, see the column 'More than nudges are needed to end the pandemic', published on 5 August 2021, by Nobel Prize–winner Richard Thaler in the *New York Times*.

Throughout this presentation, various psychological deviations will be encountered, including misperceptions, limited attention, beliefs, reference dependence, present bias, mental accounting, default option and representativeness, and we will analyse their consequences across different areas of PF, including taxation, unemployment insurance, health insurance, retirement and education. Rather than an extended survey of each topic, this Element is structured as a guide through the introduction of behavioural agents in PF, from estimable models to policy conclusions. Over the course of this presentation, some seminal issues are highlighted in order to illustrate how behaviour-oriented models could help improve fiscal policy design.[2]

Introducing the toolbox of an applied behavioural economist, this Element discusses the advantages and drawbacks of the two principal approaches in empirical PF: structural models and sufficient statistics. Interpretation of individual actions as *behavioural deviations* requires a deep understanding of informational environments. To this end, the analyst can rely on a wide range of empirical material, from lab experiments to natural experiments and survey to administrative data. Recent methodological developments in public policy evaluation offer powerful tools to recover convincing evidence of behavioural biases.

This Element builds on a wide literature at the intersection of PF and BE.[3] DellaVigna (2009) provides an overview of psychological biases and Congdon et al. (2011) discuss their consequences for policy design in the main areas of PF. To complement the empirical perspective adopted in this Element, Bernheim and Taubinsky (2018) present the core behavioural public economics models and summarise the underlying welfare principles developed in Bernheim (2009) and Bernheim and Rangel (2009). Farhi and Gabaix (2020) revisit the standard theories of optimal direct and indirect taxation with behavioural agents and further include nudges as available fiscal instruments. Other insightful references on public policies and psychological deviations include McCaffery and Slemrod (2006), Kirchler and Braithwaite (2007), Diamond and Vartiainen (2012), Mullainathan et al. (2012), Chetty (2015) and Gabuthy et al. (2021).

A deeper investigation of specific areas in PF can be found in Frank (2012) and Chandra et al. (2019) for behavioural health economics, Aaron (1999) and

[2] A large literature in BE investigates the design of properly behavioural fiscal policies through nudges, among which: Thaler and Sunstein (2003); Sunstein and Thaler (2009); Sunstein (2014); OECD (2017); Farhi and Gabaix (2020); Gabuthy et al. (2021). Although the final section briefly evokes nudges, this Element mostly focuses on traditional fiscal instruments.

[3] For a refresher on PF, see Salanié (2011); Auerbach and Feldstein (2013); Gruber (2015) and Atkinson and Stiglitz (2015).

Beshears et al. (2018) for behavioural retirement policies, Schnellenbach and Schubert (2015) for behavioural public choice and Lavecchia et al. (2016) for behavioural education. Since they have already been largely discussed by an extensive literature in BE, some topics are not covered by this Element. They are mostly related to prosocial behaviours, other-regarding preferences, self-image and signalling, in particular tax avoidance and tax evasion (Andreoni et al., 1998; Slemrod, 2007; Jacquemet et al., 2020), as well as public goods (Ledyard, 1995).

This Element deals with the following questions: Why should we care about psychological deviations in PF? How can we take them into account in theoretical models? Under which conditions can we identify and estimate their magnitude? How is it possible to disentangle behavioural deviations from incomplete information? When do we need to identify precisely the type of behavioural bias? What are the resulting guidelines for policymakers? Should individual biases be corrected?

After a presentation of the original model of optimal income taxation developed by James Mirrlees, Section 2 introduces the sufficient statistics approach to welfare analysis in PF. On top of a simpler derivation of PF models, this methodology enables the expression of optimal taxes and transfers as a function of estimable 'sufficient statistics', among which elasticities play a central role in capturing behavioural responses to fiscal incentives.

In traditional PF, these elasticities are estimated under strong assumptions regarding individual rationality and may therefore not be invariant to the way policies are framed. In order to explain the discrepancy between observed behavioural responses and their predicted theoretical counterpart, PF models started featuring behavioural agents who misperceive their environment or have non-standard preferences. Such behavioural deviations have first-order welfare effects and alter individual responses to fiscal reforms. These theoretical developments, presented in Section 3, ground a basis for the following empirical investigation.

Section 4 brings the theory of PF with behavioural agents to data. This section starts with a discussion of the conditions for the identification of these behavioural deviations. In order to be relevant for public finances, the analysis of informed choices from behavioural agents should take into account the conditions of the choice (*frames*) within public systems. As evidenced by Saez (2009), subjects do not respond in the same way to a subsidy depending on whether it is framed as a matching contribution or as a tax credit. Behavioural deviations can be revealed through changes in frames or mistakes, and should not possibly be rationalised by a standard economic model. In order to carry out this empirical step, this section presents a wide range of information sources,

from lab experiments to survey and administrative data, the combination of which opens up the way to promising investigations. A last subsection deals with the estimation of structural models and sufficient statistics with behavioural agents. Structural estimations foster extrapolation while reduced-form causal estimates of sufficient statistics can be directly plugged in optimal formulas. Ultimately, the choice between these two methods depends on the necessity to specify the nature of the bias.

Section 5 concludes, with a discussion of fiscal policies targeting behavioural agents. Behavioural deviations fundamentally challenge the core notion of welfare and push the social planner to take a stand on bias correction. Determination of the optimal policy is even more delicate when biases are heterogeneous and potentially correlated with individual characteristics such as earnings. Fortunately, BPF broadens the scope of public action through choice architecture.

2 Behavioural Responses in Public Finance

In traditional PF, there are two main reasons why fiscal reforms consider individual behaviours: on the one hand, the prevention of undesirable behavioural responses to taxes and benefits, which would increase costs for public finances or generate a sub-optimal allocation of resources; on the other hand, to encourage or discourage specific behaviours.

The first category consists of distortions. One of the first lessons in PF is that by altering relative prices, taxes and benefits change individual choices. For instance, a targeted sales tax on some goods raises their price, which reduces demand for them and creates a deadweight loss. The second category is related to incentives. The government subsidises goods associated with positive externalities (culture, energy retrofit, infrastructures, etc.) and tries to keep citizens away from activities associated with negative externalities (consumption of alcohol or sugary drinks, pollution, etc.). Policies such as the EITC rely on the assumption that labour income subsidies can be efficient in stimulating job-seeking behaviours.

In both cases, the government has to understand and forecast individual behavioural responses to tax reforms. This section discusses the role of behavioural responses in standard PF, starting with a reminder on the traditional model of optimal income taxation.

2.1 The Mirrlees Model of Optimal Income Taxation

Market failures prevent market mechanisms from generating a Pareto optimal allocation, because individuals neglect the positive or negative repercussions of

their actions (externalities), hold private information (information asymmetries) or would not engage in the production of a public good. In each of these situations, individual objectives are misaligned with social welfare, which provides a rationale for public intervention. For this purpose, the social planner relies on several instruments (taxes and transfers, quotas, production of public goods, market design, etc.) in order to influence or constrain individual actions towards a second-best equilibrium where the presence of market failures justifies the use of distortionary instruments.

The optimal income tax model developed by James Mirrlees (1971, 1986) lays the groundwork for the logic in public finance.[4] In this framework, the social planner levies taxes in order to redistribute income. Taxing away all earnings and giving the same lump-sum transfer to each agent would lead to a perfect redistribution level, but would also suppress incentives to work, which would drastically reduce tax revenue. Therefore, the social planner needs to set taxes in order to provide incentives for taxpayers to engage in productive activities.[5]

In the Mirrlees model, each individual is naturally endowed with a productivity level ω, which may be interpreted as the wage rate[6] they can get on the labour market. An hour of work will be more or less productive depending on this level. Given this *productivity type*, the agent chooses a work effort such that their *marginal rate of substitution* (MRS) between consumption and leisure equals their marginal gain from work. A government that has the ability to design type-specific income taxes would be able to influence work efforts through the control of this marginal gain. However, the social planner is not able to observe either individual levels of productivity or effort. If it were setting income taxes this way, the most productive agents would have an incentive to act as if they were less productive than they really are in order to pay a lower tax. This classic issue of adverse selection could strongly impact public finances.

Hence, the government faces an equity–efficiency trade-off: it tries to levy taxes in order to redistribute income between agents endowed with different productivity levels but cannot directly observe these productivity types. The core idea behind the Mirrlees model is that the social planner should design the optimal income tax schedule as a *truthful mechanism*, such that when faced with

[4] This presentation of the Mirrlees model is strongly based on chapter 4 of Salanié (2011). I urge readers interested in the proof of this model or theoretical developments in the economics of taxation to refer to this book.

[5] This framework, presented in terms of labour income taxation, naturally extends to all direct taxes and transfers depending on primary income, such as payroll taxes and means-tested benefits.

[6] Assuming that, in a competitive labour market, each worker is paid their productivity level. This productivity level captures several ideas, including social reproduction and unequal access to education (Saez and Stantcheva, 2016).

this tax schedule, each agent chooses the effort level that maximises social welfare for their productivity type.[7] Ultimately, efficiency is restored in a second-best economy with adverse selection.

Introducing notations helps develop a more precise representation of this model. Workers characterised by a productivity type ω who engage in a work effort l earn a gross income $z = \omega l$. A highly productive worker making a low effort can reach the same earnings level as a low-productivity worker who makes a high effort. Productivity types are distributed according to $F(\omega)$, with a density $f(\omega)$. Workers have standard preferences represented by utility[8] functions $u(c, z|\omega)$ conditional on their productivity types, increasing in consumption c and decreasing in work efforts (and thus in earnings z).[9] We further denote by $u_1(c, z|\omega)$ and $u_2(c, z|\omega)$ the partial derivatives of this utility function with respect to its first and second arguments.

The government cannot observe individual types ω nor individual effort levels l. It only observes pre-tax earnings z and sets taxes $T(z)$ as a function of this quantity. The derivative of this tax function $T'(z)$ is the *marginal tax rate*, which is the tax rate on an additional unit of pre-tax earnings. In a *progressive*[10] tax schedule, the marginal tax rate is a non-decreasing function of earnings.

Faced with this tax schedule, agents with productivity ω maximise their utility under the budget constraint $c \leq z - T(z)$, which gives the usual first-order condition:

$$1 - T'(z) = -\frac{u_2(c(\omega),\, z(\omega)|\omega)}{u_1(c(\omega),\, z(\omega)|\omega)} = mrs(c(\omega), z(\omega)|\omega).$$

Taxpayers choose an earnings level such that their MRS between consumption and effort is equal to the *retention rate* $1 - T'(z)$, which captures their marginal gain from effort. This condition is not sufficient, since high-productivity types would have an incentive to pretend that their productivity is lower in order to pay less tax. Hence, in line with the revelation principle, the

[7] This solution is grounded on the *revelation principle*, which reduces the set of mechanisms implementing the social welfare function to the subset of incentive-compatible direct mechanisms such that agents truthfully report their productivity types. Hence, the social planner can reduce their search to incentive-compatible tax schedules.

[8] Since the standard economic theory makes no distinction between the experienced utility and the decision utility, the utility function u refers to a general utility concept in this section. These utility concepts are introduced in Section 3.

[9] Given the individual productivity type, it is virtually the same to consider that agents choose their work effort or their earnings level. In line with Saez (2001), we assume that workers maximise utility over earnings z.

[10] A progressive tax schedule is characterised by a marginal tax rate greater than or equal to the average tax rate for each pre-tax earnings level z, which includes convex tax schedules as well as linear tax schedules with a lump-sum transfer.

social planner chooses a pair of functions $(c(\omega), z(\omega))$ consistent with the following *incentive-compatibility constraint*:

$$\forall \omega, \omega' \; u(c(\omega), z(\omega)|\omega) \geq u(c(\omega'), z(\omega')|\omega),$$

which states that any agent with productivity ω should be better off with the allocation $(c(\omega), z(\omega))$ designed for their type than with any other allocation $c(\omega'), z(\omega')$ designed for any other type ω'. Now assume that the MRS between consumption and pre-tax income $mrs(c(\omega), z(\omega)|\omega)$ is decreasing with productivity ω. Intuitively, starting from a given pre-tax income, this condition stipulates that any increase in effort will be more painful and less rewarding for a low-productivity type than for a high-productivity type. Under this Spence-Mirrlees condition – also called the *single-crossing condition* since it implies that the indifference curves of two agents with different productivity levels cross only once in the (c, z) plan – more productive agents earn higher pre-tax earnings and consume more than lower-productivity types.

As depicted in Figure 1, an optimal tax schedule generates a separating equilibrium between different productivity types. The income tax schedule shows the feasible set of allocations (c, z). For a given productivity type, indifference curves display the set of consumption-earnings combinations providing the same utility level. An indifference curve further away in the upper

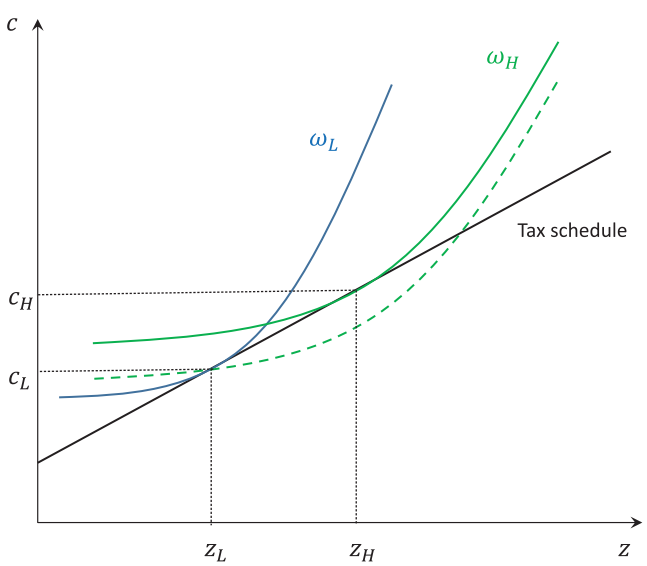

Figure 1 Separating equilibrium between two productivity types
Note: At the optimum, when faced with this tax schedule, the high-productivity type ω_H locates at a higher taxable income level z_H than the low-productivity type ω_L.

left quadrant is characterised by a higher utility level, since agents can consume a higher fraction of their earnings. A worker with productivity ω_L chooses an effort level associated with gross earnings z_L and gets a consumption level c_L. At (c_L, z_L), in line with the single-crossing condition, the MRS of a high-productivity worker characterised by $\omega_H > \omega_L$ is lower than the MRS of the low-productivity type. Hence, this worker will be better off taking the allocation (c_H, z_H) since they will be able to locate on an indifference curve characterised by a higher utility level. Under the Spence-Mirrlees condition, the tax schedule induces workers with different productivity types to choose different effort levels and consequently end up with different allocations. This separating equilibrium reveals their position in the distribution of productivity types.

Social welfare $W = \int_{\omega} G(u(c(\omega), z(\omega)|\omega))f(\omega)d\omega$ aggregates individual utilities using an increasing and concave function G that weights these utilities according to a redistributive objective. For instance, if G is the identity, the government has utilitarian preferences and maximises the unweighted sum of individual utilities. In contrast, Rawlsian preferences only consider the lowest utility level. The government is subject to a budget constraint, which states that government expenditures E cannot exceed tax revenue: $E \leq \int_z T(z)h(z)dz$, where $h(z)$ stands for the density of earnings at the optimum.

Finally, optimal income taxes maximise social welfare under the first-order conditions and the incentive-compatibility constraints for each productivity type ω as well as the budget constraint of the government. Derivation of this optimal schedule is complex, provides few insights and hardly seems related to empirical quantities. The next section presents the more recent *sufficient statistics* approach to optimal fiscal policies, which offers a nice way to overcome these issues.

2.2 The Sufficient Statistics Approach to Optimal Fiscal Policies

The *sufficient statistics* approach initiated by Piketty (1997), Saez (2001) and Chetty (2009) provides a reformulation of optimal policy instruments as a function of empirically estimable sufficient statistics. Chetty (2009) characterises this methodology as a 'bridge between structural and reduced-form methods'. Instead of designing fiscal instruments from scratch, this approach considers a small reform which is compensated by lump-sum transfers, such that the government budget remains constant. In this way, any change in instruments is directly related to welfare variations. The consequences of this small reform on tax revenue can be summarised by a *mechanical effect* and a *behavioural effect*, which represent the main forces at play in the

determination of optimal instruments. The *mechanical effect* captures variations in tax revenue assuming individual decisions remain unchanged. However, a tax reform generally alters individual incentives, inducing agents to modify their choices. The *behavioural effect* catches the impact of such adjustments on tax revenue. This distinction proves very useful when bringing tax theory to data.

Beyond direct taxation, this sufficient statistics approach has been extended to a wider range of taxes and public transfers. It proves particularly useful to evaluate marginal welfare gains from policy changes. We review here the most famous applications of sufficient statistics formulas for labour income taxation, indirect taxation and unemployment insurance. Chetty (2009) and Kleven (2021) provide an extended presentation of this methodology. Sections 3, 4 and 5 build on these formulas to introduce behavioural agents.

2.2.1 Optimal Direct Income Taxation

Saez (2001) develops a very clear and practical solution of the Mirrlees model based on the sufficient statistics approach. As shown in Section 2.1, social welfare can be given by:

$$W = \int_z \Big\{ G\Big(u(z - T(z), z|\omega)\Big) + \lambda[T(z) - E] \Big\} h(z)dz, \tag{1}$$

where the multiplier λ associated with the budget constraint of the government is the *marginal value of public funds*. Saez (2001) defines the social marginal welfare weight at earnings z as $g(z) = G'(u)u_c/\lambda$. This weight represents the utility value for the government to transfer one additional euro to an agent with earnings z. In order to simplify the derivations, we further assume no income effect.

Starting from the optimal tax schedule, consider a small increase $d\tau$ in the marginal tax rate over a small income range between z^* and $z^* + dz^*$. Such a small variation in the tax rate around the optimum should have no first-order effect on welfare but has two consequences on tax revenue:

- A *mechanical effect*: each taxpayer above z^* pays additional taxes $d\tau dz^*$. Hence, this reform mechanically increases total tax revenue by $d\tau dz^* \lambda \int_{z>z^*} h(z)dz$ but induces welfare losses $-d\tau dz^* \lambda g(z)$ for agents above earnings z^*, resulting in a total mechanical effect given by: $d\tau dz^* \lambda \int_{z>z^*} (1 - g(z))h(z)dz$.
- A *behavioural effect*: a mass $\int_{z^*}^{z^*+dz^*} h(z)dz \approx dz^* h(z^*)$ of taxpayers faced with a higher marginal tax rate in an income range between z^* and $z^* + dz^*$ have

an incentive to adjust their earnings downwards. Above this range, absent any income effect, agents do not adjust their taxable income. Around the optimal tax schedule, there are no first-order consequences of this adjustment through individual preferences, but the impact through the government budget constraint is given by: $-\lambda \int_{z^*}^{z^*+dz^*} T'(z) \frac{\partial z}{\partial(1-\tau)} \, d\tau \, h(z) \, dz$. Saez (2001) defines the elasticity of taxable income with respect to the retention rate as $\varepsilon(z) = \frac{\partial z}{\partial(1-\tau)} \frac{1-\tau}{z}$, with $\tau = T'(z)$ the marginal tax rate at z. Using approximations for dz^* small enough,[11] the behavioural effect is given by: $-\lambda\varepsilon(z^*) \frac{T'(z^*)}{1-T'(z^*)} z^* d\tau dz^* h(z^*)$.

The marginal welfare effect dW of a small variation in the income tax rate $d\tau$ is equal to the sum of the mechanical and the behavioural effects:

$$dW = \underbrace{d\tau dz^* \lambda \int_{z>z^*} \big(1-g(z)\big) h(z) dz}_{\textit{mechanical effect}} \underbrace{- \lambda\varepsilon(z^*) \frac{T'(z^*)}{1-T'(z^*)} z^* d\tau dz^* h(z^*)}_{\textit{behavioural effect}}.$$

Factoring common terms, this equation simplifies to:

$$\frac{1}{\lambda dz^*} \frac{dW}{d\tau} = \int_{z>z^*} \big(1-g(z)\big) h(z) dz - \frac{T'(z^*)}{1-T'(z^*)} \varepsilon(z^*) z^* h(z^*).$$

Equating this marginal welfare effect to zero provides the optimal tax formula presented by Saez (2001):

$$\frac{T'(z^*)}{1-T'(z^*)} = \frac{1}{\varepsilon(z^*) z^* h(z^*)} \int_{z>z^*} \big(1-g(z)\big) h(z) dz.$$

Since the right-hand side of this equation is always non-negative, marginal tax rates are always between 0 and 1. They are higher when the social value $g(z|z > z^*)$ of an additional euro transferred to taxpayers with earnings above z^* is low, when few people are concerned by this distortion ($h(z^*)$ small) and when these people are not too responsive to variations in marginal tax rates ($\varepsilon(z^*)$ small). In conclusion, empirical estimates for the elasticity of taxable income $\varepsilon(z^*)$ are crucial to evaluate the optimality of an income tax schedule.

[11] When agents adjust their earnings by dz, the non-linear tax schedule changes by $T'' = T'dz$. As explained by Saez (2001), one consequence is that the density of earnings in the optimal tax formula is a virtual density that would exist if agents were optimising with respect to the linear tax schedule tangent to the non-linear tax schedule in z^*. For a didactic purpose, these technicalities are not developed here.