Freedom Beyond Choice : Essays on the Value of Commitments in Normative Economics

Kevin Leportier

UNIVERSITÉ PARIS I PANTHÉON SORBONNE
ÉCOLE DOCTORALE D'ÉCONOMIE
Centre d'Économie de la Sorbonne

THÈSE

Pour l'obtention du titre de docteur en Sciences économiques
Présentée et soutenue publiquement
le 27 novembre 2023 par

Kevin LEPORTIER

Freedom Beyond Choice

Essays on the Value of Commitments in Normative Economics

Sous la direction de Mme Annie L. Cot
et M. Matthew Braham (Universtät Hamburg)

Membres du jury :

Mme Antoinette Baujard, Université Jean Monnet, rapporteure
Mme Constanze Binder, Erasmus University Rotterdam, rapporteure
M. Franz Dietrich, Université Paris 1 Panthéon Sorbonne
M. Marc Fleurbaey, École d'économie de Paris

L'université Paris 1 Panthéon Sorbonne n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses. Ces opinions doivent être considérées comme propres à leurs auteurs.

To Diep, who put my feet on the ground.

Contents

General introduction

What do good economic institutions achieve? We may expect them to make us happier, better, more satisfied, but also freer. In economic textbooks, one finds few occurrences of the words 'freedom' and 'liberty'[1], but whole sections and chapters are devoted to preference satisfaction and utility. Textbooks' idea of good economic institutions is that they satisfy more individual preferences than worse ones. They make people better off, according to their own idea of what is good. If we go deeper, we will also find a substantial amount of work done in economics, especially after the publication of Rawls' influential political philosophy book *A Theory of Justice* (1971), to evaluate what good economic institutions can achieve in terms of social justice, a preoccupation that has long been sidelined in standard normative economics[2]—the sub-field of economics devoted to defining criteria and methods of evaluation to apply them to economic situations. This broadening of perspective has been accompanied by a renewal of interest in the topic of freedom.

Some prominent economists have already used extensively the vocabulary of freedom in books aimed at a wider audience—Friedman and his *Capitalism and Freedom* (1962/2020), Sen's more recent *Developement as Freedom* (1999)—or more political books, such as Hayek's *The Constitution of Liberty* (1960/2011). But, with the exception of Sen, little attention has been paid, until recently, to defining precise criteria for evaluating exactly *how free* one or more individuals are in economic situations. As will be explained in the first section of this introduction, the development of a new economic literature on the measurement of freedom of choice in the last three decades brought us closer to this goal. Eventually, by bringing together these three

[1] I will use the two terms interchangeably in the following.

[2] For reviews, see Fleurbaey (1996; 2023).

different perspectives, one could hope that economists would become well equipped to address what Keynes called 'the political problem of mankind', which is 'to combine three things: Economic Efficiency, Social Justice and Individual Liberty' (Keynes 1926/2010, 311).

The point of measuring freedom, and evaluating economic situations in terms of freedom, is to appreciate how free good economic institutions can make individuals. This presupposes that having more freedom is good as such, even if it does not make us happier or serve better our preferences. The choice of economists to focus on freedom is therefore faced with obvious objections: what if people do not value having more freedom? What if, independently of what individuals think about it, having more freedom is not *valuable* in certain contexts? It is now well established in economics that in a wide range of contexts, it is better to have fewer options to choose from than more, especially: (1) when people lack the time or attention to make something valuable out of their freedom (2) when people have problems of self-control (3) when people's outcomes also depend on the actions that others take, that is, in strategic interactions. In these contexts, people are sometimes better off when they *choose not to choose*, that is, when they restrict the set of options that they will have at a later time. This is what I will call making a *commitment*.

These conclusions could be seen as pointing at the limitation of an approach centred on evaluating freedom. Not only can freedom come into conflict with our welfare and happiness, but it is also something that we would sometimes choose not to have. This thesis proposes a different perspective: what if choosing not to choose can be seen as an expression of one's freedom, rather than as its repudiation? This has already been put forward by some economists, whose positions I will present in the second section of this introduction. In any case, it is crucial to be clear about the kind of freedom that is involved in the idea that choosing not to choose can be seen as an expression of freedom. The third section will present four conceptualizations of freedom already applied to varying extents in normative economics, which may or may not provide a framework for understanding and justifying this notion.

The paradoxical nature of a choice not to choose calls for a philosophical examination. Individuals who choose not to choose experience a conflict in their preferences. If people's behaviour aligns with the standard representation of the economic agents—according to which agents have stable, consistent preferences—they would always choose what is best for them and

never feel the need for commitment since it could only worsen their situation by reducing their options. In this case, interventions interfering with individuals' choices can never make people better off. Conversely, the fact that individuals sometimes prefer not to choose implies that an intervention restricting choices might be desirable. However, this conclusion hinges on two key aspects: (1) how we represent the interests of the economic agent who commits herself and (2) our criteria of what it is desirable to achieve when individuals have conflicting preferences. The choice between different representations and criteria raises methodological questions, which I will delve into in this thesis. This is significant because it prompts economists to make substantial value judgments. It also raises ethical questions, as public interventions need to be justified by arguments destined to inform the public debate. Choices about (1) or (2) determine how convincing arguments advocating for public interventions that restrict choices can be from the standpoint of economic ethics. Section 4 of this introduction explains more precisely how these methodological and ethical questions are addressed in the thesis.

The thesis belongs to the field of philosophy of economics, which is relatively new. It is an offspring of a more encompassing discipline called philosophy of science, which analyses the methods used in the various sciences. Historically, as Hausman (2021) explains, the initial interest of some philosophers of science for economic methodology (such as Rosenberg 1976 or Blaug 1980/1992) evolved into a much broader concern for all possible intersections between economics and philosophy: both disciplines share a strong interest in rationality, causality, models, experiments, welfare and well-being, justice, the role of markets, etc. As defined by Reiss, philosophers of economics focus on the 'theoretical, methodological and ethical foundations of economics' (2013, 1). This thesis is concerned with the latter two domains: methodological, and the intersection between ethics and economics. From the perspective of economic ethics, the thesis analyses arguments that can be made to advance various public debates about the economy, with the help of economic literature and economic models. From a methodological perspective, the thesis assesses the role that value plays in normative economics as an evaluative discipline, contributing to a deeper understanding of the conceptual foundations underpinning normative economics.

0.1 Freedom in normative economics

0.1.1 Welfare and welfare economics

Economists do not only interpret the world but also aspire to transform it. A large part of economists' work and efforts are directed towards producing opinions, recommendations and advice for a decision-maker capable of implementing them, whose identity (government agency, central banker, simple citizen)—and also whose real powers and authority—often remain implicit[3]. These recommendations are necessarily based on evaluations: it is because an economic situation is in some sense 'better' than others that the economist recommends adopting the plan or policy of which this situation is the consequence. If economists stick to a positive approach—that is, describing the facts and mechanisms of an economy—they cannot provide an evaluation as such, because, due to what is sometimes called Hume's law, a factual description may not be enough to justify a value judgment identifying what is better[4]. What the positive economist can do, on the other hand, is to identify alternatives—that is, mutually exclusive policy options that are available to the decision-maker and their economic consequences. Once these alternatives have been identified, it is possible to rank them using a normative criterion that reflects values that the economist or the decision-maker holds dear. A certain set of alternatives can thus be judged better than the others. Since this latter task is—in principle—distinct from that of identifying alternatives[5], it can be assigned to a different field, which is known in economics under the name of 'normative economics'.

Historically, normative economics—the 'normative' branch of economics, which deals with evaluative statements, and not simply with descriptive ones, as does positive economics—has taken the form of what is called 'welfare economics'[6]. As the name indicates, welfare economics adopts a welfare cri-

[3]See Sugden (2018a, 17-28) for a criticism of the fiction of a benevolent social planner as the addressee of economic discourse.

[4]Hume's law or 'guillotine', which is sometimes summarized as the precept that one cannot deduce 'ought' from 'is', has been much criticized in philosophy (Searle 1964, Putnam 2002). It may not be sharp enough to establish a dichotomy between positive and normative statements. But it remains used by economists to justify the division between positive and normative economics.

[5]Chapter 1 of the thesis proposes a refutation of this claim.

[6]See Baujard (2016) for a history of the field since its beginning from the 1920s. The insistence on welfare is a product of the lasting influence of utilitarianism on the his-

terion to compare and rank alternatives. But 'welfare', as the term is used by welfare economists, means nothing more than the satisfaction of individual preferences[7]. The basic input of welfare economics' evaluations is preferences. Traditional welfare economics assume that individuals can rank alternatives according to their preferences and that this ranking is consistent and stable over time[8]. If individual preferences have these features, they can be represented by a well-defined utility function, which assigns a numerical value to each alternative: the higher this number, the more preferred the alternative. Another defining feature of preferences is that they are 'revealed' in individual choices, in the sense that any possible choice between alternatives by an individual is supposed to reflect the preference ranking of this individual[9]. The higher the alternative is in his preference ranking, according to this preference satisfaction criterion, the better off the individual is. The formal

tory of economic thought, although the definition of what counts as welfare has evolved considerably.

[7]This association defines what I will call in the following the preference-satisfaction criterion for evaluating economic situations.

[8]Another relevant aspect of preferences in the traditional welfare economics framework is that they are usually self-regarding and defined over what Sen calls *culmination outcomes*. To use an example from Sen, 'An arbitrary arrest is more than the capture and detention of someone—it is what it says, an arbitrary arrest' (Sen 2009). The kind of utility that critics of welfarism have in mind is in principle only about the culmination outcome (the capture, detention), not the comprehensive outcome, which would also include process-related aspects, such as the arbitrariness of the arrest. If people's preferences are assumed to focus only on culmination outcomes, then all process-related aspects—including the fact that the individual was able to choose from a sufficiently rich set of options—are not taken into account in the utility functions that represent these preferences.

[9]There is an important debate in philosophy of economics about the merits of this understanding of preferences (see in particular Hausman 2011, Dietrich and List 2016 for criticisms of this approach). Johanna Thoma thus defines 'revealed preference theory': it equates 'preference with actual or hypothetical choice behaviour. According to that understanding, when an economist ascribes a preference for an option a over an option b to an agent (...), this is meant to capture nothing more than that the agent does or would choose a over b from some specified choice set' (Thoma 2021a, 164). A weak preference relation is defined as transitive and complete. A choice function gives the set of alternatives chosen by an agent in each possible set of alternatives. The agents' choices, as described by a choice function satisfying Houthakker's axiom—which states that if an alternative a is chosen when b is available, then whenever b is chosen, a is also chosen when available—, can be represented by a preference relation (see Kreps 1988). Houthakker's axiom is the the minimal condition to be imposed on the structure of choices in order to be able to say that choices reveal preferences.

framework of revealed preference is well-defined in economics, but it leaves unanswered a crucial normative question: why does preference satisfaction matter so much when it comes to public policy?

As McQuillin and Sugden (2012) explain, the formal framework of welfare economics had the merits of enabling to conduct broad evaluations of public policies (through judgements about efficiency and cost-benefit analysis) while leaving open the philosophical interpretation of the criterion, which makes the framework, despite its seemingly narrow focus on preferences, welcoming to different and conflicting normative views. Sugden and McQuillin enumerate three different interpretations that have been given of the criterion, and which all could justify its adoption:

1. The utilitarian perspective of early neoclassical economists such as Pigou relied on the possibility of utility measurement as a cardinal and interpersonally comparable quantity, where the utility was understood as referring to a sort of 'hedonic experience'. In modern, post-paretian, welfare economics, preference satisfaction can be taken as an (ordinal) index for measuring levels of individual happiness: the more utility someone has, the happier they are. This *happiness interpretation* would thus bring the maximal satisfaction of everyone's preferences close to Bentham's 'greatest happiness of the greatest number'.

2. according to the *well-being interpretation*, preferences are indicators of what an individual 'judges to be his well-being, or of what he is trying to achieve' (McQuillin and Sugden 2012, 555). The closer an individual is to his most preferred alternative, the better off he is, according to his own standards, and that would be a good thing for anyone identifying well-being with the satisfaction of subjective preferences (McQuillin and Sugden 2012, 555).

3. according to the *freedom interpretation* (or consumer sovereignty interpretation), under a conception of revealed preferences, preferences can be seen as a way of summarizing the choices that individuals *would* make, under appropriate circumstances. Since economists take preference rankings as given without making any judgements about them, individuals are left free to choose what they really want. As this interpretation concerns freedom, I will come back to it later.

The problem with this broad perspective on welfare is that when 'preferences cannot be assumed to be coherent, these different normative positions

have divergent implications' (ibid., 555). If, as suggested by the findings of behavioural economists showing how inconsistencies in choice behaviour are systematic and prevalent[10], preferences (as revealed by choices) cannot be assumed to be consistent, it becomes unclear what should constitute the basis for evaluating, respectively, what makes an individual happy, what promotes their well-being (as they see it), what they would choose among a certain set of alternatives[11]. Roughly, advocates of the happiness interpretation will try to define other measures of happiness, often based on self-assessment surveys[12], while proponents of the well-being interpretation will try to figure out what are the 'true preferences' of the individuals, which 'makes them better off, as judged by themselves' (Sunstein and Thaler 2003)—true preferences being the preferences that individuals would reveal in their choices if they were not affected by biases, mistakes, or problems of self-control. Advocates of a freedom or consumer sovereignty criterion have tried to preserve as much as possible the traditional approach by concentrating on choices that robustly reveal preferences—under a vast array of 'ancillary conditions' (this is the program of Bernheim and Rangel 2009)—or to define a completely different criterion, based on the reference to 'opportunities' rather than welfare (Sugden 2018a)[13].

In any case, economists find themselves confronted with a 'reconciliation problem'—the 'problem of how to reconcile normative and behavioural economics' (McQuillin and Sugden 2012, 554). The shallow consensus that up-

[10]See Sugden (2018a, 7-13) for a presentation of how these findings shake the foundations of welfare economics by questioning its traditional behavioural hypotheses. An example of such inconsistencies is the endowment effect, that is, 'the fact that people often demand much more to give up an object than they would be willing to pay to acquire it' (Kahneman et al. 1991, 194), whereas people endowed with consistent preferences would give the same monetary value to the object whatever the context.

[11]Saint Paul (2011) shows, in a similar perspective, how the behavioural turn in economics has resulted in a change of perspective on welfare, which, in particular, has the consequence of making paternalistic intervention relevant.

[12]See Frey and Stutzer (2002) for a review of the field which is now called 'economics of happiness'.

[13]The literature devoted to amending welfare economics to take into account the lessons of behavioural economics—without abandoning the preference-satisfaction criterion—can be called 'behavioural welfare economics', following Bernheim and Rangel (2009). Since Sugden's proposal, it may be classified under the label 'behavioural normative economics'. In the following, I will use the term 'normative economics' to refer to the 'normative' branch of economics under all its manifestation, which includes the various proposals for using and defining criteria other than the preference-satisfaction criterion.

held traditional welfare economics as a workable approach no longer exists, as most economists now take seriously the findings of behavioural economics[14]. As McQuillin and Sugden note, 'many substantive questions in moral and political philosophy' that 'did not need to be asked' because of this consensus are now coming to the surface. New, ambitious proposals, such as Sugden's promotion of an 'opportunity criterion' and his program of a 'normative economics without preferences' (Sugden 2021), gained some traction in the field of normative economics. In his 2018 book, Sugden criticises the new 'behavioural welfare economics' for its reliance on a 'preference purification' approach which is not grounded on any received psychological theory. According to Sugden, it is purely gratuitous to assume that there exists a set of consistent 'true preferences' lying behind the inconsistent choices that individuals make as if an 'inner rational agent' was to be found inside the actual, irrational individual (Infante, Lecouteux and Sugden 2016). Sugden's alternative 'opportunity' criterion appeals to the value of freedom:

> I use the term opportunity in the sense that is standard in economics and social choice theory: an opportunity for an individual is something that he has the power to bring about, if he so chooses. Some writers (...) use the unqualified term 'freedom' for this concept; Sen (1995) calls it 'effective freedom'. However, the proper definition of freedom is a contested issue in philosophy; for my purpose, it is simpler and more transparent to speak about opportunity.[15] (Sugden 2010, 49)

Using an opportunity criterion has the merits, in Sugden's view, of not requiring any assumption about preferences to conduct evaluations—individuals who have opportunities are simply free to choose whatever they prefer at the moment of choosing. A motivation for introducing a freedom criterion is therefore that it would permit addressing the reconciliation problem.

These debates and developments of the last two or three decades would be impossible to understand, however, if we did not mention the traditional reluctance of economists to make value commitments. This reluctance explains why economists do not usually endorse any one of the interpretations

[14]See Angner (2019).

[15]In the following, I will use the term 'opportunity' in this exact sense, and relate it to freedom in the same way.

of the preference satisfaction criterion. It can also explain why economists accept the preference-satisfaction criterion: because it enables economists to leave the task of making value judgements to the individuals themselves. This position is characterized by Haybron and Alexandrova as 'normative minimalism':

> Normative minimalism is a set of implicit principles of welfare economics. It purports to keep value commitments to a minimum, if not to avoid them altogether, notably by orienting normative economics solely towards the satisfaction of preferences, and thus (ostensibly) deferring to individuals' own value judgments. (Haybron and Alexandrova 2013, 159)

Interestingly, normative minimalists would thus naturally adopt an anti-paternalist stance—according to which it is for each individual to judge what matters for them[16]. Anti-paternalism is usually motivated by a concern for values such as individual autonomy or equality. Here, the reluctance to make any value commitment morph into anti-paternalist deference to individuals' own value judgments[17], which make the preference satisfaction criterion appealing to economists. But the behavioural turn in economics has also shaken the shallow consensus on anti-paternalism, in addition to welfare economics itself: while new measures of happiness have a paternalistic flavour[18], some of those who wants to keep track of individuals' own idea of their well-being call themselves 'libertarian paternalists'[19] because, like Thaler and Sunstein, they propose to alter individuals' environment to influence their choices while leaving their freedom of choice unaffected. The new 'behavioural paternalism', which can also be found in philosophy (Conly 2013), generated a backlash from some economists and philosophers[20], which brought the issue of what can philosophically justify paternalism to the fore—making it relevant for economics to ask why paternalism should or should not be opposed.

It would not be accurate, however, to say that philosophical questions about the normative criterion of welfare economics were not asked and answered before the advent of behavioural economics. Amartya Sen's writings

[16] On anti-paternalism and welfare economics, see Sugden (2018a, viii).

[17] which is as such a form of value commitment: see chapter 4, section 1 on this issue.

[18] As shown, in particular, in McQuillin and Sugden (2012).

[19] See Thaler and Sunstein (2008).

[20] Among the economists see: Sugden (2018a), Rizzo and Whitman (2020), Saint-Paul (2011).

can be credited for giving rise to a large movement of interdisciplinary work devoted to defining alternative ways of evaluating economic situations. Sen's starting point as a normative economist is Arrow's impossibility result establishing the impossibility of defining a social welfare function aggregating individual preferences into a single collective preference, under seemingly reasonable conditions[21] (Arrow 1951/2012). This result seemed to make it impossible to define a procedure for evaluating economic situations which would be based on individual preferences, and not on some 'dictatorial' judgement about what is good. One of the diagnostics made by Sen about this impossibility result was that it was due to artificial informational poverty (Sen 1979). In particular, restricting oneself to only accept ordinal preferences as inputs of the aggregation procedure would not be a defensible position for economists concerned with the evaluation of economic situations (whose distributive aspects are clearly politically and morally relevant but cannot be assessed without interpersonal comparisons of utility). This led Sen to extend his criticism to welfare economics' choice of normative criterion. The exclusive focus on preference satisfaction, which Sen called 'welfarism'[22], imposes, according to Sen, 'severe constraints on the type of information that may be used in making social welfare judgements', because it excludes 'non-utility information', that is, any kind of information about individual choices and welfare which is not expressible *via* a ranking reflecting individual preferences. Non-utility information matters:

> There are principles of social judgement that require essential use of non-utility information, and while such principles (e.g. liberty, non-exploitation, non-discrimination) are typically not much discussed in traditional welfare economics, they do relate closely to the subject matter of welfare economics (Sen 1979, 547)

A second motivation for introducing a freedom criterion is thus the recognition of the informational limitations of welfarism, which led to Arrow's impossibility theorem. Even if this impossibility result could be circum-

[21] Arrow showed that if such a function satisfies the condition of Pareto optimality, unrestricted domain and independence of irrelevant alternatives, it is necessary dictatorial.

[22] Welfarism can be defined as the principle according to which 'social welfare is a function of personal utility levels, so that any two social states must be ranked entirely on the basis of personal utilities in the respective states (irrespective of the non-utility features of the states)' (Sen 1979, 538).

vented by dropping the requirement that utility should be ordinal and non-interpersonally comparable, much was left outside the evaluative scope of normative economics. Sen had already proposed, in his famous 1970 article about 'The Impossibility of a Paretian liberal' to integrate non-utility information into the arrovian framework of social choice theory. But the article delivered an impossible result: it demonstrated the incompatibility between the Pareto principle and a condition of 'minimal liberty' which was inspired by John Stuart Mill. This condition stated that there should exist some alternatives over which an individual is sovereign, in the sense that the collective preference over these alternatives always reflects the preference of this individual—and their preference alone. The idea is the following: if the only difference between one alternative and another is that the inner walls of my house are painted yellow in one case, and green in the other, then only my preferences should be taken into account when it comes to deciding whether one alternative or another should be preferred collectively. This condition, while formulated in the welfarist Arrovian framework, required information about the delimitation of what Sen called the 'protected sphere' of an individual—the set of personal matters that should concern no one but this individual.

Sen showed that this condition of minimal liberty could sometimes conflict with the Pareto principle. His 1970 article launched a vigorous debate, which is still alive today. But Sen did not stop there. In his Tanner lectures, published in 1980, he proposed, in answering the question 'equality of what?', to replace the utilitarian concern for utility or the Rawlsian concern for resources or primary goods as the *equalisandum* of a theory of justice, by a concern for what he called 'capabilities', defined as the freedom to achieve valuable 'doings' and 'beings'. Ultimately, Sen can be credited for launching three different strands of research regarding freedom:

1. *Rights.* What Sen called 'minimal liberty' was later interpreted, by Gibbard (1974) in particular, as a right[23]. Gibbard (1974) showed that, within a slightly different framework than Sen, the simple fact

[23]One important difference between a freedom and a right is that an individual's right 'implies obligations on the part of others agents to do or not to certain things, but an individual's freedom does not necessarily imply any such obligations of others' (Pattanaik and Xu 2009). A thief is free to steal unattended property, but we would not say that others have the obligation not to interfere with the stealing. However, according to certain conceptions, having one's rights enforced and respected *is* what it takes to be a free person. I will come back to this in section 3.

of granting such individual rights to at least two persons could entail incompatibilities, without even mentioning the Pareto principle. A literature, which is still alive today, has explored in much detail these incompatibilities. Since its focus is more on the compatibility between rights (and the Pareto principle) than on providing normative economics with different normative criteria, I will not have much to say about its accomplishments. But it generated important conceptual discussions about the right way to model rights. Nozick (1974, 164-166) objected to Sen that individual rights should not be defined *through* Arrow's aggregation procedure to form collective preferences (by imposing a condition of minimal liberty on the procedure), but as a way of fixing certain 'features of the world' when exercised (*prior* to any aggregation procedure). Several theorists proposed that, instead of conceiving rights as a relationship between individual preferences and collective preferences, they should be represented as a relationship between possible actions that an individual can take (if they so choose) and the outcomes of a social process[24]. The latter representation of rights illustrated what Sen called later a conception of 'liberty as control' (Sen 1985), according to which there can be liberty or freedom[25] only when individuals are in control of the outcomes. Sen defended his own approach by advocating a new concept of 'indirect liberty', to which I will come back in section 2.

2. *Capabilities.* Sen's efforts to 'unstrap the straitjacket of preferences' (Anderson 2001) finally led him to define a completely different approach to the evaluation of an individual's well-being, called the capability approach. In this approach, what matters from the point of view of evaluation is not utility or preferences but 'functionings' defined as the 'doings' and 'beings' (for example, being well nourished, being in good health) that people value in their lives. The 'capability' of an individual is defined as the set of all mutually exclusive functioning bundles which are available to that individual (Pattanaik and Xu 2020). Thus, the capability of an individual reflects her freedom to choose among alternative functioning bundles. The term 'capability' can also simply refer to 'what people are able to be and to do' or to

[24]see Gärdenfors (1981), Gaertner et al. (1992), Sugden (1985) for justifications of these conceptual choices and criticism of Sen's approach.

[25]As many, in the following I will treat these two words as having the same meaning.

put it differently, 'a person real *freedoms* or *opportunities* to achieve functionings' (Robeyns 2017, 39). Indeed, Sen assimilates the capabilities to achieve functionings to what he calls 'real freedoms' (Sen 1995, 149). But as Pattanaik and Xu note, we can doubt that the notion of capability as real freedom captures everything valuable or relevant about freedom. Freedom is often contrasted with *ability* (Miller 1983), in particular among libertarians. Suppose that the government prevents me from taking a course on Marx's theory of value: this limits my freedom. And yet, this restriction would not count as a restriction of my capabilities if I could not really follow the course anyway, for example because I am too dumb to understand it. Those who contrast freedom with ability would say that I lost the freedom to understand Marx's theory of value even if I did not have to ability to do it. But Sen's definition of capabilities as real freedoms fails to capture this distinction. The thesis will therefore also consider other approaches to evaluate freedom.

3. *Freedom of choice.* The capability approach can be used to evaluate the justice of certain distributional arrangements, as an answer to the question 'equality of what?' The extent of the *advantage* that a situation gives to someone over others can be evaluated in terms of capabilities. We could, following Sen's suggestion, compare the capability set of someone with that of others to see how much more he can do or be than others. The wider the set, the freer an individual is to choose among functioning bundles. This makes it necessary to find approaches to measure capability sets and compare them[26]. A pioneering article by Pattanaik and Xu (1990) formed the starting point of a new literature devoted to defining rules for rankings sets of options, called 'opportunity sets' (sometimes also called 'menus'), to determine whether some set would give more freedom of choice than another to the person choosing. While the nature of the sets remains open to interpretation in this literature—which makes it more relevant to evaluate freedom as such than the literature on capabilities—, opportunity sets can readily be interpreted as capability sets. The next section will explain in more detail the contributions and limitations of this literature (henceforth called 'The freedom of choice literature').

[26]Sen (1999a, 33-45) proposed rules to compare capability sets, but did not take a firm stance on the right way of doing it.

0.1.2 The freedom of choice literature and the independent value of freedom

The basic input of any welfare analysis, as it is conducted in traditional welfare economics, is the ranking of alternatives in terms of individual preferences. Here is how Arrow presents the framework:

> We assume that there is a basic set of alternatives which could conceivably be presented to the chooser. In the theory of consumer choice, each alternative would be a commodity bundle; in the theory of the firm, each alternative would be a complete decision on all inputs and outputs; in welfare economics, each alternative would be a distribution of commodities and labor requirements. (...) These alternatives are mutually exclusive; they are denoted by the small letters x, y, z,... On any given occasion, the chooser has available to him a subset S of all possible alternatives, and he is required to choose one out of this set. (Arrow 2012, 12)

Ranking *sets* of alternatives S, T (or 'opportunity sets') is an exercise different from ranking alternatives x, y, z. From a welfarist perspective, all that matters is that individuals get their most preferred alternatives. We can thus derive a *indirect-utility rule* for ranking opportunity sets from this perspective, which states that the value of a set is the value of its best alternative, with respect to individual preferences. In other words, a set S is better than a set T, according to this ranking rule, if and only if the best alternatives in S are better than the best alternatives in T. According to this rule, the value of an opportunity set can never be increased by adding suboptimal alternatives to the set, nor can it be decreased by removing suboptimal alternatives. Such changes are completely indifferent: only the best alternatives matter.

Intuitively, this indirect utility rule does not reflect any concern for freedom. Given equilibrium prices, an individual buying goods in a competitive market is free to choose any bundle of goods that they can afford. But suppose, as suggested by Pattanaik and Xu (1990)—the two authors of the seminal article that launched the freedom of choice literature—that our market economy is replaced by a command system led by all-knowing benevolent bureaucrats. If the bureaucrats force individuals to consume their most preferred bundles under equilibrium prices, individuals' preferences are just as

satisfied as they are under the market system—and therefore the two opportunity sets associated with each system are just as good according to the indirect utility rule—and yet, it would seem, they have no freedom. If people, in the absence of uncertainty, would (strictly) prefer to live in a market economy rather than in this system, it would show that freedom matters to them *as such*, and not simply because it enables them to better satisfy their preferences. They care about having suboptimal options—options that they would never choose—in addition to their most preferred ones. I will say that, in the case where the indirect utility rule is contradicted in this fashion, individuals value freedom *independently* (of its ability to satisfy preferences), or that they give freedom an independent value.

The question thus becomes: what ranking rule could capture the independent value of freedom? The freedom of choice literature tries to answer this question by analysing different ranking rules and showing how they can be mathematically characterized in terms of formal axioms, which are meant (when they are not deemed to be purely technical) to capture various normative aspects associated with freedom[27]. Consider for example the 'simple cardinality-based' rule for ranking set, characterized by Pattanaik and Xu (1990). This rule simply ranks opportunity sets by counting the number of alternatives they contain: three alternatives are always better than two, whatever may be the content of the alternatives. Pattanaik and Xu show that we would need to adopt this rule if and only if we accept the three following axioms:

- *Axiom of indifference between no-choice situations:* the ranking rule should be indifferent between all opportunity sets that contain only one alternative.

- *Axiom of strict monotonicity:* the ranking rule should rank the opportunity sets that contain two alternatives higher than those that contain only one.

[27]I will not try to distinguish the notion of 'freedom of choice' from that of 'freedom' (see Carter 2004, Oppenheim 2004 for attempts to do it). As shown by Baujard, 'it is clearly evident that there is no unity in the freedom of choice literature. In particular, it can be seen that the axioms do not necessarily belong to a single concept of freedom and this is a source of confusion' (Baujard 2007, 248). See section 3 below for a discussion of the relevance of definitions of freedom.

- *Axiom of independence:* adding the same option to two sets should not change the way these two sets are ranked.

This simple cardinality rule is a very crude measure of freedom of choice, which completely disregards the value or quality of the alternatives for the person choosing—everything is just the same. But the axioms that characterize the rules might have a certain plausibility for someone who values freedom. Thus, Pattanaik and Xu's characterization reads like an impossibility result. Something is wrong with at least one of these axioms, which needs to be amended to allow the consideration of less crude ranking rules. This impossibility result emulated the efforts of economists and philosophers to find a workable approach to measuring freedom. Two influential objections were raised against Pattanaik and Xu's axioms:

- *concern for the value of alternatives.* Sen (1991) disputed the idea that adding an alternative to an opportunity set should always be seen as an improvement in freedom (a position which, in particular, involves rejecting the axiom of strict monotonicity). Offering someone the additional possibility of being beheaded at dawn, to use Sen's example, would not improve in any way his freedom since it has no value to him—or anyone who is not suicidal. The thrust of Sen's argument was that evaluations of opportunity sets in terms of freedom of choice cannot be conducted without taking into account the value that individuals give to the alternatives that the set contains.

- *concern for the diversity of alternatives.* Another issue, which was raised by Pattanaik and Xu in the initial paper, is that adding the same alternative to two different opportunity sets does not necessarily increase freedom as much in both cases (thus contradicting the axiom of independence). Suppose that you take a blue-coloured bus to go to work. You could also, equivalently, take the tramway. Being proposed to take the tramway *in addition* to take the blue-coloured bus would, arguably, offer you more freedom than being proposed to take an otherwise identical red-coloured bus in addition to the blue coloured bus. The red-coloured bus is not significantly different from what you already have, and would not add as much to your freedom as the possibility of taking the tramway. Evaluations in terms of freedom would thus need to take into account how diverse the alternatives of a given set are.

While connected, the concern for diversity differs from the concern for value in one major respect[28]: the value that an alternative has does not depend on the set which contains it, but an alternative is distinguishable or different only in relation to other alternatives in the set. I will not survey the contributions to the literature, which often attempt to respond to one or the other of these two strands of criticism[29]. What I want to examine here is a central claim, common to most ranking rules that are defined and discussed in the literature, which is that it is always possible to add some alternatives to any opportunity set to make it better with respect to freedom. This is obvious in the case of the simple cardinality rule: adding any alternative would make the set better. Rules that address the concern for the value of alternatives would qualify this assertion: only alternatives that have enough value, or provide enough utility, make it better. As for the concern for diversity, only alternatives that add enough diversity to the set make it better[30]. This general claim may be plausible when it comes to evaluating freedom: arguably, one can always have more freedom, just as one can always be better off or have more utility. But we may doubt that individuals would always benefit from having more freedom in this sense—would it really be better for them, *on the whole*, to always have more relevant alternatives to choose from?

Let us consider the crucial presupposition at the heart of Pattanaik and Xu's approach: that individuals care or should care about being free to choose alternatives—even suboptimal alternatives that they would never choose. Why should they bother having suboptimal alternatives, rather than just go with their most preferred one? This might, after all, seem a bit perplexing and needs justification[31]. Surveys of the literature put forward two main kinds of justification[32]:

- *global well-being.* The first kind of justification is Millian in tone and

[28]These aspects had already been discussed by philosophers before economists, in an informal manner: see Gray (1990, 33) for a review.

[29]Barbera et al. (2004) is a very detailed survey of the literature on ranking sets. For surveys more focused on the freedom of choice literature, see Pattanaik and Xu (2015), Dowding and van Hees (2009).

[30]In the following, I will speak of 'significant alternative' to designate an alternative that contributes to the freedom of individuals either because it is valuable enough or because it brings enough diversity.

[31]As will be detailed in the next section, experimental evidence suggests that individuals often do not value having choice as such.

[32]see Barberà et al. (2004, 926), Pattanaik and Xu (2015, 366).

appeals to the notion of what we might call 'global well-being' or 'utility in the largest sense'[33] which would include a concern for welfare in the strict welfarist sense and a concern for freedom[34]. In the words of Sugden (2006), 'Even if, as moral observers, we were confident that we knew which way of life it would be best for some individual to choose, we would still promote his well-being most effectively by letting him choose for himself, making his own mistakes and learning from them'. According to Mill, the 'free development of individuality is one of the leading essentials of well-being' (Mill 1859/ 2006), but individuality cannot 'grow' if the individual does not exercise and develop his own capacity to choose[35]. This implies that we cannot promote global well-being without individuals being provided with a sufficiently rich array of alternatives from which they can choose. This first justification is therefore instrumental, although it gives freedom an independent value[36].

- *autonomy.* The second kind of justification is Kantian in tone and appeals to the notion of autonomy, which, in the Kantian tradition, is not reducible to global well-being. Its modern version asserts that an autonomous person is someone capable of reflecting on the preferences

[33]'I regard utility as the ultimate appeal on all ethical questions, but it must be utility in the largest sense; grounded on the permanent interests of man as a progressive being' (Mill 1859/2006, 17).

[34]This raises the question of how much weight to give to freedom and utility. See Baujard (2011) for a survey of the literature on pluralist rankings reflecting several values.

[35]'The human faculties of perception, judgement, discriminative feeling, mental activity, and even moral preference, are exercised only in making a choice' (Mill 1859/ 2006)

[36]Ian Carter (1995) makes an important distinction between the *independent* value of freedom and its *intrinsic* value. I will formulate this distinction as follows: freedom has an *independent* value for someone when they sometimes strictly prefer having suboptimal alternatives in their opportunity set rather than not. If this is not the case, it means that the value of any alternative for someone only *depends* on their contribution to their utility level. Freedom has an *instrumental* value for someone if having more freedom is (empirically) correlated with having some other thing (for example, global well-being) that this person values more and pursues. The crucial implication is that someone can give freedom an independent value (as does Mill) and at the same time value it instrumentally, because all that matters, eventually, is another value like global well-being. Freedom has an *intrinsic* value for someone if it is not valued instrumentally. If someone values freedom because it gives them more responsibility, or enables them to better express their autonomous self, they value freedom intrinsically because the connection between freedom and responsibility or autonomy is conceptual and not empirical.

they would like to have *and* to act according to these preferences[37]. Freedom is valued, in this perspective, not because of its effect on well-being, but because it is a genuine expression of someone's autonomous self. This would also imply that individuals should be provided with a rich array of alternatives because it enables them to better express their autonomous selves. The idea is the following: the choice that someone would make among several alternatives is more indicative of his autonomous preferences than if he could only get what he prefers the most. Consider the case of buying a present for one of your loved ones: the more alternative you have, the more expressive your choice will be of whatever feeling you want to convey to the recipient of the gift. One obvious downside is that more choice involve more responsibility, but this is precisely what an autonomous agent shall aspire to in a Kantian perspective[38]. This justification, unlike the previous one, gives freedom an intrinsic value[39].

What these two perspectives highlight is that the value of an opportunity set for an individual may depend also on suboptimal alternatives that are never chosen. According to the Millian justification, rejecting alternatives is formative of someone's individuality and makes us learn and grow. According to the Kantian justification, rejecting alternatives is what makes an autonomous choice expressive of one's autonomous preferences. This gives a clear answer to the question of why an alternative we would never choose should matter. But it is very noticeable that endorsing one or the other of these justifications would lead to holding different views of what counts as meaningful freedom, although the literature rarely attempted to connect justifications and ranking rules (or axioms)[40]. More importantly, these justifications do not imply that it would always be better to have *more* alter-

[37]See Binder (2021a) for a definition along these lines.

[38]See van Hees (2022) for an argument along these lines.

[39]Scanlon (2000, 254-255) enumerates three ways in which choice can be valued: it can have (1) a 'predictive' value, because what people choose would be more likely to satisfy them, (2) a 'representative' value, which is connected to what I called the Kantian justification ('I want [my choices] to result from and hence reflect my own taste, imagination, and power of discrimination and analysis'), and (3) a 'symbolic' value, because having people choose on your behalf on important matters is *demeaning*. This third value of choice, while important because it is connected to the idea of equality in standing, is not often considered by economists, which is why I will not say much about it. See also Raz (1988).

[40]See, however, Baujard 2007.

natives to choose from, or even that it is always good to be able to choose. We cannot infer that having choices is always good, from the perspective of global well-being or autonomy.

These justifications are only *conditional*. From the Millian perspective, choices are a good exercise of our faculties of choice only when they offer someone the opportunity to grow and learn from experience. From the Kantian perspective, choices are valuable only when they enable someone to choose in ways that are expressive of their autonomous selves. But it cannot be presumed that every choice we make offers such an opportunity. When these situations of choice satisfy the conditions for the Kantian and Millian justification to have some bite, the ranking rules defined in the freedom of choice literature may be used to evaluate the degree of freedom that someone enjoys. But what if that is not the case? The next section will present general contexts where the Kantian and Millian perspectives would not justify that we should have more alternatives to choose from. This leaves us with an important question: what should we think of the contexts, or situations of choice, where neither the promotion of general well-being nor the expression of someone's autonomous self, justifies giving special importance to freedom? Should we give up on the idea that using a freedom criterion can be appropriate when evaluating circumstances such as these?

0.1.3 Choosing not to choose

The behavioural turn in economics, and especially in normative economics, is a call to take into account the actual decision-making process of individuals when trying to identify their preferences and values—rather than admitting that their behaviour would necessarily satisfy the usual rationality assumptions. Behavioural economics seeks to increase the explanatory power of economic theory by integrating inconsistent or biased choice behaviours into economic models. Welfare behavioural economics enlarges the scope of welfare economics by enabling it to make judgements about the *welfare* of inconsistent economic agents. In a somewhat similar spirit, the goal of this thesis is to discuss and suggest ways to make normative judgements about the *freedom* of economic agents when they are faced with situations of choice which do not respond to the standard justifications—presented in the previous section—for using a criterion of freedom of choice. The objective is to enlarge the scope of normative economics, by pursuing the long-standing effort to provide welfare economics with a freedom criterion—to be able to

apply such a criterion even when the standard justifications for freedom fail to apply.

The Kantian and Millian perspectives on freedom, which justify the project of evaluating situations in terms of freedom of choice, make the case that freedom *should* be valued independently—because of its importance for global well-being or autonomy—, whether or not individuals agree with it. But what if people actually do *not* value freedom independently? This would put economists committed to anti-paternalism *and* to producing evaluations in terms of freedom in an awkward position[41]. If people do not value freedom independently, evaluation in terms of freedom would not capture something that individuals value, and that would contradict the anti-paternalist stance that individuals should be the judge of what matters to them[42].

How would we know that individuals value freedom independently? They can welcome having more alternatives to choose from even if freedom is not valuable as such for them. Individuals who only care about the satisfaction of their preferences (defined over alternatives) would value having larger opportunity sets to choose from if they are unsure of what will be their preferences at the time of choice—this is what Kreps (1979) called a 'preference for flexibility'. Therefore, the observation that people value *larger* sets would not be a good indication that they value freedom independently. However, the observation that people value choosing in *smaller* sets would be difficult to reconcile with the view that they value freedom independently, especially if they are ready to get rid of significant alternatives. Le Lec and Tarroux (2019) conducted an experiment on such 'choice-aversion', by eliciting individuals' monetary valuations of opportunity sets containing consumption items. One of their main results is that, on the aggregate, the monetary valuation of an opportunity set is significantly lower than the value of its preferred element—this difference is used to measure the extent of choice aversion. Furthermore, one-half of the subjects are on average choice averse

[41]This point was already made by Fleurbaey: 'People not only have preferences over ordinary dimensions of their lives, but also about the amount of choosing that they have to go through. Respecting their view of the good life includes taking account of their attitude toward the size of the menu. Adopting an exclusive focus on opportunities is unlikely to be respectful in general.' (Fleurbaey 2012, 438-439)

[42]Haybron and Alexandrova (2013) proposed the notion of 'inattentive paternalism' to describe how cost-benefit analysis evacuate (in their view) some value commitments that individuals might have from the evaluation. Chapter 4 is devoted to exploring this important issue in much more detail.

in this sense, and only between one-quarter and one-third of them have a preference for choice. Le Lec and Tarroux thus conclude about the value of freedom that 'the premise that the size of the choice set is an important component of well-being' suffers from 'a lack of support'.

A clarification is needed at this point. The term 'choice aversion' is ambiguous. There is a crucial distinction to be made between 'not choosing' and 'choosing not to choose', as emphasized by Sunstein (2015). The phenomenon of 'choice overload', made famous, among other things, by Barry Schwartz's popular book *The Paradox of Choice* (2016) comes to the mind of many when the discussion about the value of choice arises. The paradox of choice would be that it renders us neither happy nor free, because of choice overload. Choice paralyzes. But also, according to Schwartz, choice 'tyrannizes' (Schwartz 2016, 2). In the section of his book devoted to the phenomenon, Sugden defines choice overload as a situation where 'Consumers face so many options that the quality of their decisions declines, or they feel dissatisfaction with their final choices, or their motivation is so undermined that they would avoid choosing altogether' (Sugden 2018a, 143). Choice overload is often interpreted by economists as implying that people fall back on the default alternative when the size of the opportunity set increases[43]. This is what we may call 'not choosing'. But, as such, the fact that individuals are not choosing says nothing about whether or not individuals value freedom independently because it says nothing about their attitudes towards opportunity sets.

Choosing not to choose is different from not choosing, which is pure abstention[44]. The choice not to choose is voluntary and dynamic in nature: it means that an individual is ready at time t_0 to *restrict* the opportunity set

[43]Iyengar and Lepper (2000)'s seminal paper on choice overload implemented a setup where a range of high-quality jams where presented on a table to people shopping in a supermarket, with a one-dollar discount offered if a jam was bought. On 'limited' choice days, six jams were presented. On 'extensive' choice days, twenty-four jams were offered. People were less likely to buy a jam on extensive choice days (two percent did) than on limited choice days (12 percent did). Interpreting not buying as the default alternative, this result suggests that individuals' choices are inconsistent since they would change their behaviour and fall back on the default alternative when more alternatives were added.

[44]In the setting which is described here, not choosing *is* choosing because the default is an alternative among others. From the Kantian perspective, good-doers are acting responsibly only when doing evil—or less dramatically, omitting to do good—is permissible. Similarly, doing nothing when you could do something is a potential omission and mistake that the Millian perspective would allow you to make to develop your individuality.

that they will have at an ulterior time t_1. The examples are many: locking all the alcohol you own in a cupboard and throwing away the key, hiring someone to decorate your apartment, burning your bridges in a battle, etc. There are many reasons why individuals would—or even should—choose not to choose, which I will present in the rest of this section. Some of these reasons also explain why individuals would fall prey to the phenomenon of choice overload. But only the choice not to choose (which I also call a *commitment*) is indicative of the fact that individuals do not value freedom independently, because they are ready to get rid of significant alternatives. In this case, one can say that they have a *preference for commitment*[45]. If the economists do not defer to this second-order preference, we may say that they are guilty of 'choice-requiring paternalism', to borrow Sunstein's expression[46]. Choice-requiring paternalism would force individuals to choose or, more subtly, disregard their preference for commitment[47]. By contrast, we may, following Sunstein, call 'forced active choosing' the fact that people are prevented from avoiding choosing (by falling back on the default alternative). Choice-requiring would force people to choose in larger sets than what they would prefer, while forced active choosing would simply remove the default alternative. These distinctions are reported for clarity in Table 1.

Whether or not people value freedom independently, there are some contexts where the Millian and Kantian perspectives do not give any justification for doing so—such as when individuals have limited attention, self-control problems, or face externalities. As I will argue, in these contexts choices would not express one's autonomous self, or enable someone to exercise their faculties of choice to develop their individuality. Therefore, in these contexts, a choice not to choose is not something that would be incompatible with valuing freedom independently. I also have selected these three contexts because they are already well-known in economics and because I will refer to

[45]The expression is taken from Gul and Pesendorfer (2001), which define what we may call 'second-order preferences' on sets, rather than on alternatives.

[46]On this concept, see Sunstein (2015,113-153) and chapter 4.

[47]As Sunstein recognizes, the term 'choice-requiring paternalism' is highly paradoxical. How could it be paternalist to give *more* alternatives to choose from to individuals? As explained in chapter 3, an intervention counts as paternalistic if it takes steps to make it more difficult or impossible for people to choose what they prefer, with the intention of making them (and them only) better off. If an individual would choose not to choose, but some third party prevents it from happening for the sake of this individual's freedom, we have a case of paternalism according to this definition.

	Representation in models	*Associated form of paternalism*	*Possible motivations*
Choosing not to choose	choice of a smaller opportunity set	choice-requiring paternalism	limited attention, self-control problems, externalities
Not choosing	choice of the default alternative	forced active choosing	choice overload

Table 1: Choosing not to choose

them in the chapters of the thesis:

- *Limited attention or time.* It is sometimes argued that the cognitive costs involved in the effort of identifying the alternatives and evaluating them can explain or justify a preference for commitment[48]. But what to make of the casual observation that one can always restrict one's thinking or attention to a limited subset of a (large) opportunity set (as Schwartz himself notes[49])? When the opportunity set expands, we can always ignore the new alternatives. One possible answer is that disregarding the additional alternatives is in itself an act of will, which may have a cost[50]. The problem with choices in larger sets would thus lie in the need to exert willpower to direct one's attention to the task ahead (or alternatively, and more simply, the time to do it), which is considered a scarce resource by some psychologists. Duflo (2012) interprets Banerjee and Mullainathan's (2008) model of poverty trap as implying that mandates could enable people to allocate their limited

[48]Barry Schwartz (2016) makes this point. Ortoleva (2013) provides a representation theorem for the preferences of an individual who suffers from 'thinking aversion' and, as a result, dislikes larger sets.

[49]'Why can't people just ignore many or some of the options, and treat a 30-option array as if it were a 6-option array?'

[50]According to some psychologists, there is a cognitive cost to be paid to *avoid thinking* about irrelevant items or aspect of the context. Duflo (2012) cites Wegner et al.'s (1987) experience, where people are instructed to list their thoughts and some are, at the same time, asked to avoid thinking about a white bear. People who were also asked to avoid thinking about a white bear listed fewer thoughts than the others.

attention to more rewarding tasks and increase their productivity. In the model, individuals have to divide their limited attention between problem-solving at home or at work and have to choose how much 'comfort goods' they buy—which makes problems at home less costly. The model implies that public interventions aimed at limiting the occurrence of problems at home—and thereby limiting the range of choices that need to be made to solve these problems—would enable people to pay more attention at work, making them more productive. As Duflo summarizes, 'The problem here is not that individuals make the wrong choice on the home front because they just go with the flow. It is that the energy and time they expand making the right decision on the home front takes away from the other things they could be doing with their time' (Duflo 2012, 19). Put simply, the problem with having more freedom of choice in a certain area of life is that some choices are a distraction. Since time, attention or the emotional energy needed to exert willpower is limited, consuming more of these scarce resources in activities that individuals' autonomous selves value less is not a good exercise of their autonomy. Similarly, the more latitude a person has in dividing their attention between activities according to his values and tastes, the more likely he is to develop his individuality and character. The more trivial the choices, the less relevant the Kantian and Millian perspectives become to justify the importance of choosing. In rich countries, notes Duflo, citizens have paradoxically much less to choose when it comes to the 'basic constituents of life'[51], and yet 'They are likely to be largely on the right track' because the right choices are already made for them. It would make no sense to provide citizens of rich countries with untreated water in addition to the clean water to which they are used—even if this would make them face the same larger opportunity sets as citizens of poorer countries. Not because individuals would not value freedom independently, but simply because these choices are a distraction to most people[52].

[51] 'The richer you are, the less responsibility you need to take for the basic constituents of your life (retirement savings, clear water, immunizations).' (Duflo 2012, 3)

[52] 'Paternalism, far from being opposed to individual responsibility, may form a basis on which we might have freedom over what really matters in life. most of the choices the poor have to make are just pure "noise", which at best stands in the way of them making important choices and at worst leads them to make a wrong turn and fail to achieve the basic amenities needed for a decent life' (Duflo 2012, 23)

- *Self-control.* Following Schelling's work on the 'non-self-governing consumers', who 'behave sometimes as if they had two selves, one who wants clean lungs and long life and another who adores tobacco, or one who wants a lean body and another who wants dessert' (Schelling 1978, 290), behavioural economists have developed models of 'dual selves'[53], which explains why individuals sometimes have a preference for commitment. In particular, models inspired by Thaler and Shefrin (1981) describe the potential conflict between a long-run 'planner' self whose plans maximize the lifetime utility of the individual and a short-run 'doer' self who makes consumption decisions but focuses on the present. This setup makes it clear that the planner is taken to be the 'authentic' self of the individual, whose 'true preferences' (as Thaler and Sunstein 2008 put it) are the only one that matters normatively. In such a situation, expanding the opportunity set of the doer would not make much sense from either the Kantian or Millian perspective. Because the preferences of the doer for immediate gratification conflict with the preference of the authentic self, it prevents the latter from expressing themselves and acting responsibly. Similarly, it would not be very formative for the planner's individuality to let the doer thwart his plans—from the point of view of the planner, this would be an inconvenience, not a useful mistake. Because we have identified the interest of the person with that of the planner self, the choices of the doer self only matter insofar as they serve those interests. There is therefore no justification for valuing the doer's freedom independently.

- *Strategic interactions and negative externalities.* Individuals would often benefit, in contexts of strategic interactions, from having some of their freedom to act limited. Not because these choices are a distraction, but because other people get in the way of achieving valuable outcomes if we have choices. Strategic interactions and negative externalities, just as self-control problems, prevent individuals from carrying out their plans. But by contrast with the two latter cases, in such contexts, individuals may be perfectly rational and still have a preference for commitment because sometimes, as Dixit and Nalebuff put it, freedom to choose is simply freedom to lose (Dixit and Nalebuff 2008, 168). I will present three such cases, which will be considered in the thesis:

[53] These models are presented in more details in chapter 5.

– *Public goods.* According to the public good argument already made two centuries ago by Adam Smith, individuals would not contribute enough to financing public infrastructures if they were left to decide on their own of their contribution, because of assurance or free riding problems[54], even if the public good is of immense value to them. A public intervention forcing everyone to contribute the optimal amount would thus make everyone better off, from their own point of view. However, no preference for commitment is possible in this setting: people would simply not accept getting rid of their freedom to contribute as they wish—at least as long as a sufficient number of other individuals do not commit to doing the same, in certain circumstances (as explained in chapter 2).

– *Credibility problems.* People sometimes have an incentive *ex ante* to promise or threaten other individuals to do things that, *ex post*, they will not do, because carrying out the promise or the threat *after* individuals have complied or rebelled brings no reward at all, only costs. Knowing this, these other individuals would thus never give in to the empty threat or promise that has been issued. A monopolist would be better off deterring a potential competitor by threatening to launch a price war. But the monopolist also has an incentive not to launch it once the competitor has entered the market, which makes the threat empty. Schelling's solution to this credibility problem is simply to publicly get rid of the possibility of *not* carrying out the promise or the threat[55]: this move is usually called a 'strategic commitment'.

– *Markets and negative externalities.* The opening of a new market gives everyone new opportunities for exchanging and acquiring goods they can afford. But new markets can also create negative externalities which affect how people fare and make some individual or collective outcomes impossible to achieve. According to textbook models of externalities, firms would not stop at the optimum and produce too many goods, failing to generate an efficient level of externalities in the absence of corrective mechanisms. Peo-

[54]See chapter 2 and section 4 of this introduction for a more precise formulation.

[55]See Schelling (1960/1980) for a more detailed presentation, through numerous examples. See also chapter 3 for an analysis.

> ple may well collectively refuse the freedom to perform these new transactions if they are deprived of the possibility of achieving other valuable outcomes as a result of opening such a market.
>
> There is no clear justification to draw from the Kantian or Millian perspective to expand individuals' freedom to act in these cases if it makes them unable to carry out their plan—not because of their mistakes or faults, but because of the obstacles that others put in their way. The argument here is very similar to the one given above: *ex ante*, the agent, in the position of the planner, would want to commit to using a certain strategy to carry out his plan—financing the public good, carrying out a threat or promise, producing the optimal quantity of goods—, but fails to do it *ex post*, in the position of the doer: because it is better *ex post* to free ride, or to go back on one's promise. In all these cases, even if individuals do not suffer from self-control problem, the nature of social interactions ensures that having more choice of strategies undermine the individuals' ability to carry out their plans. It thus prevents the expression of the individuals' autonomous preferences without enabling them to learn or develop their faculties of choice.

The choice situations that I just presented are now very standard in economics. In these contexts, a preference for commitment may have some relevance even for someone who cares about global well-being or autonomy (and would thus value freedom independently in normal contexts)[56]. But the freedom of choice literature has not much to say about it, as it does not usually consider commitments and preferences defined over opportunity sets. More than a gap in the literature, it is a frontier. It would need to be crossed if evaluations in terms of freedom should lay claim to the title of alternatives to welfarism, or at least to supplement the traditional welfare economics approach. What could reasonably be said about freedom in these cases?

[56]Dold and Lewis have recently developed a different but related argument based on Sen's distinction between opportunity and process aspect of freedom. They claim that 'more opportunity may be associated with a reduced sense of agency and a person may prefer to have fewer options if the increase in her sense of agency offsets the loss in opportunities' (Dold and Lewis, 2023, 4).

0.2 Normative economists on commitments

I presented at the end of the last section cases where having more significant alternatives to choose from is not better, even for those who value the Millian development of individuality or autonomy. I also explained why, in these cases, individuals may be expected to have a preference for commitment. Where does this leave us? We could react either by saying that this simply shows the limitations of the attempts to formulate a freedom criterion to evaluate economic situations and the weakness of the reason for doing it[57], or we could follow the steps of renowned economists with philosophical inclinations (such as Sen[58], Schelling, Sugden, Buchanan) and try to make sense of commitments from the perspective of a freedom evaluation. This is what the thesis intends to do. This section will present existing works in normative economics which consider preferences for commitment from the point of view of freedom. There is no identifiable literature either in normative economics or in philosophy devoted to that topic[59]. To my knowledge, Sen and Sugden are the only economists who have described precise criteria which acknowledge commitments as such. Notions of 'strategic commitment' and 'preferences for commitment' are well-known to economists, and widely

[57]If providing individuals with large opportunity sets is more costly than just giving them what they would choose in these sets then one can make the case, as Fleurbaey suggests, that this is not only paternalistic but *wasteful*: 'A society that spends resources to guarantee a wide menu of opportunities to each citizen is wasteful if everyone would rather be given a narrower menu that better fits their preferences' (Fleurbaey 2012, 438).

[58]Note that I am not using the term 'commitment' in the sense that Sen (1977) gave it to designate one of the two possible foundations for other-regarding behaviour: 'We must distinguish between two separate concepts: (i) sympathy and (ii) commitment. The former corresponds to the case in which the concern for others directly affects one's own welfare. If the knowledge of torture of others makes you sick, it is a case of sympathy; if it does not make you feel personally worse off, but you think it is wrong and you are ready to do something to stop it, it is a case of commitment' (Sen 1977, 326). Commitment, for Sen, involves a 'counterpreferential choice' (ibid., 328), while sympathy does not. But it has nothing to do, in Sen's terminology, with the act of restricting one's options.

[59]In philosophy, Mill's refusal of slavery contracts—an extreme instance of a choice not to choose—has triggered debates about the interpretation of Mill's harm principle; this will be tackled in chapter 5. A problem connected but different from commitments is that of the alienability or inalienability of rights. In economics, there have been some discussions about how to formalize alienable rights in the wake of the discussion of Sen's Paretian liberal (the discussion starts with Gibbard 1974). Philosophers have also sometimes discussed how to measure freedom *across* lifetime (Carter 2011, Schmidt 2017), which is a very complex and more general problem.

used and applied. But little effort has been made to inquire about their normative significance from the point of view of freedom. Preferences for commitment have a very paradoxical character, which makes it a challenge to assess their value from this point of view. But, as this section intends to show, they are key to understanding three debates about the economy which have wide implications about what an economy should look like, and will be important themes explored in the thesis.

0.2.1 The choice architecture of the modern world

'Choice architecture' is a somewhat vague term used by Thaler and Sunstein (2008) to indicate 'the background against which choices are made'. It is inevitable, according to them, that (at least some) choices are influenced by background conditions and economists should therefore pay attention to them instead of assuming that only the content of the alternatives available influence decisions. I propose to extend the term to include not only conditions determining *how* and *what* we choose, but also what we *can* choose. Norms, regulations, and public infrastructures determine in this sense the choice architecture of a modern, rich society because they provide us with some alternatives and exclude others. Duflo's example of water coming out of the tap clean in rich countries is illustrative of how public infrastructures determine what we choose, since, as she says, 'it would take some work to figure out how to opt out of drinking clean, treated water' (Duflo 2012, 8-9) and choosing to drink clean water or not is not really a choice to be made. Duflo's observation, in the wake of Thaler and Sunstein[60], is that in rich countries people are 'on the right track' because the choice architecture made it so, even when norms or public infrastructures do not leave us any choice but the right one. Extrapolating a little, we could say that the problem is not, as Schwartz (2016) and contemporary critics of the consumer society assume, that, because of the pressures of capitalism, we have presently too many choices compared to the past but that we are *still* not steered enough toward the choices that matter to us.

Interestingly, for Duflo[61], the fact that our range of choices is sometimes reduced with the provision of public infrastructures can be counted as an

[60]'Whatever Mill might have thought paternalistic interferences with freedom of choice are hardly absent from nations that generally respect liberty' (Sunstein, 2015, 14).

[61]For an extensive criticism of Duflo's 'democratic paternalism', see Favereau (2021), chapter 9.

increase in freedom: 'I am in fact more free in a society that puts chlorine in my water even if I did not explicitly ask for it than in a society that does not' (Duflo 2012, 16). This echoes what Sen said about 'freedom and disadvantageous choices'. Choice is disadvantageous if it 'forces on the person the necessity to spend time and effort in making lots of choices that he or she would rather not have to make' (Sen 1995, 62), which means that the existence of such a choice may generate a preference for commitment. Instead of a conflict between 'freedom' and 'advantage', what we have here is rather a conflict between what Sen calls different types of freedom[62], 'The freedom to exercise active choice over a range of (possibly trivial) options and the freedom to lead a leisured life without the nuisance of constantly having to make trivial choices'. And since 'The expansion of some types of choice can reduce our ability to choose life-styles that we might treasure' (Sen 1995, 63), we might even think that less choice would sometimes lead to more freedom. But the reasoning that lies behind this judgement is not very clear. There is no clear argument for weighing these different dimensions of freedom, as Sen presents them—range of choices versus ability to choose lifestyles—in a way that entails Duflo's paradoxical conclusion that less choice produces more freedom.

Fortunately, Sen offers a way forward. He gives a few pages later another way of conceptualizing freedom, which stems from the recognition that choice, in addition to being sometimes disadvantageous, is also often *impossible* in modern society:

> given the complex nature of social organization, it is often very hard, if not impossible, to have a system that gives each person all the levers of control over her own life. But the fact that others might exercise control does not imply that there is no further issue regarding the freedom of the person; it does make a difference in how the controls are, in fact, exercised. (Sen 1995, 65)

As emphasized by Sen, the fact that it is relatively safe to walk in the streets of most modern cities, or that crippling diseases like polio have been eradicated is a huge accomplishment of rich, modern societies, opening up a

[62]See Fleurbaey (2012, 439) for a similar point: 'By giving people what they want, including as regards menus size, one grants them a more valuable form of freedom than by forcing them to have a menu of formidable size.'

host of opportunities for everyone, unimaginable in a less scientifically advanced and policed environment. But the possibility to walk safely or be unaffected by contagious diseases is not something that, properly speaking, we choose. And it is neither something that we *can* choose because these outcomes are the product of complex processes of decision, coordination, and standardization, which no one can bring about or even plan on their own. The fact that we, as end users, do not even have a say in these complex processes may even be inevitable to produce the desired outcome. Telling people to experiment by themselves whether we should all drive on the left or the right side of the road, for example, would not produce the outcome that everybody wants. It can only happen if the choice to drive on the left or the right is somehow dictated: coordination works better when we have no control over its implementation. The choice architecture of modern societies makes certain choices impossible or irrelevant, to produce good outcomes that would never exist without these extensive—but often unnoticed—restrictions. According to Sen, even if we cannot really choose these outcomes, their realization is something that, under good conditions, we *would* choose. Therein lies the connection to freedom.

> As long as the levers of control are systematically exercised in line with what I would choose and for that exact reason, my 'effective freedom' is uncompromised, though my 'freedom as control' may be limited or absent. (Sen 1995, 65)

This conception of 'effective freedom', or 'indirect liberty', as Sen calls it elsewhere (Sen 1982), might be exactly what we need to evaluate the choice architecture of the modern world in terms of freedom. The conception of freedom implicit in the freedom of choice literature refers to an agent pulling the strings, and more or less constrained in his ability to do so[63]. And yet, sometimes we delegate certain choices to others (as when we do proxy voting): we inform them of our preferences and leave them control over the realization of outcomes. In other cases, as Sen remarks, we do not even delegate anything but as we cannot be in control of the decision, people, knowing our preferences, act to make things happen as we would choose it. This is the case when someone bleeding and unconscious receives first aid:

[63]See Sugden's definition of opportunity given in section 1. Barberà et al. define an opportunity set as 'the set of all feasible (mutually exclusive) options from which the agent can have any option by simply choosing to have it' (Barberà et al. 2004, 924).

it is assumed that the person would accept the treatment that is applied to him to save his life. Suppose he would not: in this case, Sen would say that his 'indirect liberty' is not served. However, that does not mean that we can always make decisions on behalf of others as if it did not matter whether the person actually has the lever of decision in his hands. Sen is cautious to add that control matters for freedom. We may even think that 'indirect liberty' can only be relevant when someone is no longer in control of choices that matter to them[64]. In any case, Sen's notion of 'indirect liberty' extends the relevance of freedom to contexts where the conception of freedom received in normative economics would have nothing to say.

This offers us a way to think about freedom in situations where choices are a nuisance, a distraction, or block the realization of more important outcomes. Sometimes we are well in control of our choices, but the second-order decision to choose to choose—or not to choose—is not in our hands, because we simply cannot make a commitment, just as Duflo's poor. If possible, we may want to transfer our power to decide to someone if we could be sure that they would choose just as *we* would. But suppose that, for some reason, we cannot. Then, Sen's argument for extending the relevance of freedom applies also here: our 'indirect liberty' is well served when the government sets up public infrastructures that bring about the outcomes that we would have chosen. This extension of Sen's idea of 'indirect liberty' is used and discussed in the second chapter of the thesis.

0.2.2 Markets, freedom and commodification

Debates about the value and role of the market have a long history. Critics of the market have notably claimed that (exclusive) reliance on markets to supply goods fosters inequality, and poverty, undermine communities and the value of solidarity, or lead to various spectacular market failures—among which the destruction of biodiversity or climate warning. More recently, some influential works on philosophy (Anderson 1990, Radin 2001, Satz 2010, Sandel 2012) came up with a different strand of criticism, which bears on what we may call 'market attitudes'. As Sandel puts it, 'Markets don't only allocate goods; they also express and promote certain attitudes toward the goods being exchanged' (Sandel 2012, 9). Therefore, the extension of

[64] An important task is indeed to specify what kind of circumstance may justify an intervention by which we act on someone's behalf without their prior consent. The task is left for us, as Sen did not elaborate on that. See chapter 2 on this issue.

the sphere of the market, which we have, according to Sandel[65], observed since at least the end of the Cold War, is also the extension of these market attitudes which tend, according to critics, to replace other attitudes which form the very fabric of a good society. We can try to convince children of the intrinsic value of learning, or we can pay them to read books. We can appeal to citizens' altruism to donate blood to provide transfusions to those in need, or we can simply pay them to do it. In the latter cases, we have what has been called 'commodification' (the act of 'allowing certain things to be for sale' Brennan and Jaworksi 2016, 19).

Critics of the market put forward many arguments. I will consider these two:

- *crowding out argument.* The first, which relies on an empirical premise, is that the existence of a market for these goods or services tends to 'crowd out' the intrinsic motivation of individuals to provide them[66]. This could lead to a very special case of market failure, where the supply of a good or service is reduced (or its quality altered) when it becomes tradable. This argument has attracted the attention of economists and there is now good evidence that the empirical premise of this argument is sometimes correct[67]. Those who provide a good or service for free, out of intrinsic motivations, might not do it anymore for a payment, which reduces its supply. They might not even continue to provide it for free anymore, because once a price is put to, say, a pint of blood, 'my giving a pint of blood is like giving fifty dollars of my money' (Radin 2001, 96), and that is not the kind of gift I was willing to make.

- *semiotic argument.* The second argument relates to freedom and is purely analytical. It asserts that the simple *possibility* of being paid to

[65]'As the cold war ended, markets and market thinking enjoyed unrivalled prestige, understandably so. (...) Even as growing numbers of countries around the world embraced market mechanisms in the operation of their economies, something else was happening. Market values were coming to play a greater and greater role in social life. Economics was becoming an imperial domain. Today, the logic of buying and selling no longer applies to material goods alone but increasingly governs the whole of life. It is time to ask whether we want to live this way.' (Sandel 2012, 6)

[66]See Bowles (2016) on the question whether changing economic incentives actually work, which is not discussed here

[67]There is now extensive literature on this question, see for example Gneezy and Rustichini's (2000) famous paper on Haifa day-care centres, evocatively entitled 'A fine is a price'.

do an act which was previously done out of intrinsic motivations, alters the very significance of this act[68]. As Radin puts it: 'We cannot know the price of something and know at the same time that it is priceless. Once something has a price, money must be part of the interaction, and the reason or explanation for the interaction' (Radin 2001, 101). Therefore, the argument goes, commodification deprives everyone of the freedom to give 'priceless' things, or to establish with someone a non-market relationship of giving. This argument is a form of what has been called the 'semiotic argument' by Brennan and Jaworski (2016).

This latter argument was originally formulated by Titmuss (1970) to include unfreedom among the harmful effects of a market for blood transfusions. Commodification deprives individuals of the freedom to enter into a 'gift relationship'. It was later criticized by Arrow (1972); Arrow's reasoning was simply that commodification makes it possible for everyone to sell *and* to give, and thus enhances everyone's freedom of choice compared to the situation where it was only possible to give. This terse answer—which I call the 'simple argument for markets' is representative of many economists' vision of the connection between freedom and markets. Because commodification gives everyone direct access to something they value, provided they can afford it, without (seemingly) removing anything from people who do not want to engage in these monetary transactions, it only *adds* opportunities. As Fabienne Peter vividly puts it, 'markets appear not only as mechanisms that efficiently allocate resources but—beyond that—as systems that automatically legitimize themselves' (2004, 3). The fact that market allocations are efficient is not trivial, since markets may well be prone to failure. But the fact that they 'automatically' bring new opportunities to the table would be enough to create a prima facie presumption in favour of the extension of the sphere of the market. Amartya Sen, who often made the case that markets promote freedom[69], has been careful to distinguish two sides of the argument: the fact that the market gives us automatically more *freedom to act* does not necessarily mean that it gives us more *freedom to achieve*[70], since the latter depends on the capacity of the market mechanism to provide valuable outcomes to everyone—which is not the case if, for instance, the externalities

[68]This argument was made, among others, by Titmuss (1970), Singer (1973), Anderson (1990), Radin (2001).

[69]See chapter 1 for references.

[70]A distinction made in Sen (1994).

that result from market activities worsen, on the whole, many individuals' lives and make it impossible to reach certain achievements.

Amartya Sen seems to endorse Arrow's reasoning, once qualified and limited to the idea that commodification brings everyone more freedom to act. But what if commodification changes the significance of the act to give itself, as Anderson and Raddin suggest? Economists have not been very attentive to the way market opportunities are value-laden, even if this is the reason why commodification sometimes generates strong reactions from the public. One notable exception to this observation is Alvin Roth, who characterizes 'repugnance' to some transactions as 'constraints on markets'. As Roth remarks, 'Attitudes about the repugnance (or other kinds of inappropriateness) of transactions shape whole markets and therefore shape what choices people face' (Roth 2007, 38). If phenomena of repugnance are constraints that individuals impose on the opening of new markets, they are also constraints that individuals impose on themselves, because they could transact and exchange these 'repugnant' goods but choose not to. This idea, combined with Arrow's argument, suggests that shared repugnance could be interpreted as a form of collective commitment.

However, in light of the philosophical debate about commodification, the use of the term 'repugnance' may not be ideal, since it seems to identify opposition to commodification as a mere—and contingent—distaste, when for some philosophers, commodification changes the *nature* of the goods, which would justify reasoned opposition to commodification. This significant gap between economists' and philosophers' conceptualisation of the same object (market opportunities) calls for an examination of how economists conceive what we may call 'normative models' (Sugden 2003, Beck and Jahn 2021) to evaluate economic situations. Such normative models are based on a certain description of opportunities, which, as Sugden argues, involve certain value judgements. The purpose of the first chapter of this thesis is to elucidate these judgements and show how Arrow's simple argument in favour of commodification can be resisted, by reinterpreting the 'semiotic argument' in terms more familiar to economists. The analysis also shows that opposition to commodification cannot be read as a form of commitment, since, contrary to what Arrow said, some significant freedoms to act are sometimes lost as a result of the opening of a market.[71].

[71]People make a commitment, as defined above, when they choose in t_0 to choose in a *smaller* set in t_1. But refusing the opening of a market cannot be modelled as choosing to

0.2.3 Capitalism, addiction and self-liberation

Historian of drugs and critic of 'limbic capitalism' David Courtwright describes addiction as a 'compulsive, regret-filled pursuit of transient pleasures that are harmful to both the individual and society', a 'very bad habit, in the sense of being strong, preoccupying, and damaging, both to oneself and to others' (Courtwright 2019). This description, which corresponds neither to the medical nor the economic definition—which is, in the framework of Becker and Murphy (1988), solely based on the nature of the addict's consumption path—reflects the use of the term in current and media discourse. In addition to tobacco, drugs or alcohol, or 'substance' addictions, one can be addicted to gambling, sex, watching porn, gaming, working, shopping, eating, exercising, scrolling, etc.—the list of 'behavioural addictions' being seemingly infinite[72]. Addiction in this sense is seen by individuals as time-consuming, hard to resist, and they regret engaging in it because it prevents them from following their higher goals. But 'limbic capitalism', as Courtwright puts it, which embarks technology in its endeavour to 'engineer, produce, and market potentially addictive products in way calculated to increase demand and maximize profit', thwart these aspirations by making addictive goods or activities *accessible*, *affordable*, conspicuous (via *advertizing*), and its consumption *anonym* and *anomic*. This environment would make any attempt to exert self-discipline and manage one's addictions much more difficult. Unavoidably, this discourse ends with a call to regulation: to regain their freedom, citizens need to tame the beast and prevent entrepreneurs from exploiting their limbic-system-induced weaknesses.

A lot of normative assumptions remain undiscussed here. The model of human agency implicit in this discourse, as Tyler Cowen (1991) remarks, is that of a division between a 'rule-oriented self', and an 'impulsive self'—the 'desirability of victory for the rule-oriented self', which attempts to constrain the behaviour of the other, being presupposed, but rarely defended as such. We could on the contrary, as Tyler Cowen does, emphasize the multiple ways in which individuals could suffer from excessive discipline. The 'impulsive self', argues Cowen, is the bearer of values like spontaneity, self-discovery, and risk-taking. Too much self-constraint may thus result in making someone frustrated, overly rigid, and incapable of spontaneity. In this regard, capi-

choose in a smaller set if some significant alternatives are lost in the process. See chapter 1 for a more detailed analysis.

[72] A justification for this extension is provided by Ainslie (1992).

talism, with its marketing and advertising techniques, fosters self-liberation precisely because it erodes the overdiscipline of the 'rule-oriented self', and, hopefully, enable the individual to strike a better compromise between the aspirations of both selves. This reversal of perspective is enlightening but does not tell us much about how to evaluate the freedom, or unfreedom, that modern capitalism brings about when it alters and interferes with individuals' self-management—especially since capitalism can also make a profit out of preferences for commitment. As Ainslie (1992) remarks, since addicts both want and do not want the addictive good, they may be willing to buy it and at the same time to buy an antidote for it—an irrationality that firms can exploit. One notable development is indeed the rise of markets for products that play the role of commitment devices, often under the name of 'productivity apps' or 'focus apps', which can block access to certain other applications during a certain time or condition the failure to reach a certain objective to a punishment ('write or die' is the suggestive name of one of these applications) which is contractually agreed upon *ex ante*. From a regulatory point of view, the question thus becomes to know how the state should regulate this market. What should be the extent of the punishment that private companies are contractually allowed to inflict on users who broke their commitment? Should they be large, to promote self-discipline, or kept limited, to promote self-liberation?

In his article 'Ethics, Law and the Exercise of Self-Command' (1985), Schelling already explored this question from an evaluative point of view: 'If somebody now wants our help in constraining his later behavior against his own wishes at that later time, how do we decide which side we are on?' Schelling rejected both the idea that it is possible to identify an 'authentic self' (as critics of capitalism siding with the 'rule-oriented self' would tend to do), and that some kind of intra-personal comparison of utility between selves is possible or feasible, which could determine which commitment represent the best trade-offs between selves. Schelling also notes that 'full freedom entails the freedom to bind oneself, to incur obligations, to reduce one's range of choice. Specifically, this is freedom of contract'. But freedom of contract is the freedom to contract with someone *else*, under the understanding that one will receive something, at some point, in exchange for the reduction of one's range of choice. There is no contract with oneself, only a *vow*, remarks Schelling. Could the state enforce a simple vow? And to what extent? Schelling explores in his paper the consequences of this legal innovation but does not determine whether it should really be done and how. In any case,

Schelling's rejection of the assumption of preference stability opens the way to further elaborations, as it raises the central normative issue that needs to be addressed: what should be done if there is no clear argument for taking one side rather than another in intra-personal conflicts?

James Buchanan, who also questioned the assumption of preference stability, provided some answers. In 'Natural and Artifactual Man' (1979/1999), he advocated a 'strong defense of individual liberty' which values it independently, in contrast to modern welfare economics, influenced by its utilitarian heritage which Buchanan rejects. The model of intertemporal preferences, which represents the individual as 'maximizing the present value of his utility stream', does not leave room for persons who 'imagine themselves to be other than they are' and who 'take actions designed to achieve imagined states of being' (Buchanan 1999, 253). For example, by choosing to commit oneself to not smoking, a smoker not only prevents his preference for smoking from being satisfied, but also changes the kind of person that he is. 'His preferences shift; he becomes the non-smoker that he had imagined himself capable of becoming'. The logical conclusion is that, as the individual has to choose between 'imagined futures', he 'remains necessarily uncertain as to how that which he chooses will work out. He has a clear interest in seeing that the choice set, the set of alternative imagined futures, remains as open as is naturally possible, and, if constrained, that the constraints be also of his own choosing' (ibid., 259). Any attempt to forcibly close off his future options would thus harm his interests. Can we conclude that Buchanan would support giving a free rein to a market for commitment devices that inflict material penalties to the users, in a fully libertarian spirit? What is perplexing is that Buchanan insists that, because the individual 'does not, and cannot, predict that person he may want to become in subsequent period', he wants to 'keep his options open'. But why would he choose to constrain his own future behaviour if it is so important to him to keep his options open? There is an unresolved tension at the heart of Buchanan's conception of freedom[73].

The more advanced attempt to define a freedom criterion for evaluating situations where preferences are dynamically unstable is that of Robert Sugden[74]. While Sugden shares with Buchanan his commitment to the prin-

[73]Lewis and Dold (2020) explore these issues in more detail.

[74]Sugden's opportunity criterion, extended to evaluate multi-periods, is described in Sugden (2006; 2007; 2018a)

ciple that 'liberty should be understood in terms of what [someone] is free to choose' which implies that, in the absence of preference stability, opportunity sets should be as open as possible, he rejects what he calls Buchanan's 'ethic of self-creation' (Sugden 2018b, 28) and the centrality of the effort to shape one's preference to justify this principle. Freedom or opportunity is valuable even for those who are not trying to become anything else than what they are. What justifies the use of a freedom criterion is simply that preferences are likely to change and that individuals have an interest in seeing their preferences satisfied, whatever they may turn out to be. Sugden's opportunity criterion is grounded on the idea of a 'responsible agent', who treats her past and future actions as her own, 'even if she does not yet know what they will be, and whether or not she expects them to be what she now desires them to be' (Sugden 2018a, 106). But this would seem to rule out the possibility that a responsible agent should use commitment devices, as it would mean that she does not really treat their future action as her own (as Schubert 2015 and Fumagalli 2023 remarked).

Sugden's opportunity criterion, when applied to the evaluation of multi-period decision problems, remains curiously neutral on this question. A commitment device can be modelled as an opportunity given at time t_0 to close some later opportunity available at time t_1 (that is, an *opportunity for commitment*). I will speak of a *hard* commitment device in this case, or in the case where a material penalty is attached to the choice of this opportunity. A *soft* commitment device, by contrast, would not close this opportunity but make it less likely that it is chosen by attaching to its choice a psychological penalty (an example of a soft commitment device is a self-nudge, as Reitjula and Hertwig 2022 describe them)[75]. According to Sugden criterion[76], opportunities to use hard commitment devices—are neither freedom-enhancing nor restricting. They simply add nothing of substance. This position has been criticized by Schubert (2015) whose own proposal, a criterion of 'opportunity to learn', seems closer to Buchanan's position in that he values the possibility of using (hard) commitment devices if having too much choice hinders the process of learning an forming new preferences. In chapter 5, I formulate a different perspective, which I borrow from Mill's famous argument against slavery contracts. It requires neither endorsement of the postulate of the responsible agent nor an ethics of self-creation, but concludes that allowing

[75]See chapter 5 for more details on this distinction.

[76]The exact definition of the criterion is presented in chapter 4.

hard commitment devices is something that a 'liberty principle' such as Mill's cannot do, because it would undermine the very purpose of the principle.

Sen and Sugden provide original and valuable frameworks for considering commitments from a freedom perspective. Sen explains, thanks to his notion of indirect liberty, how individuals' freedom can be paradoxically 'served' even when they have no alternatives to choose from. But his proposal remains sketchy, as he never tries to specify the right circumstances under which it would be acceptable to take away people's control over certain outcomes to better serve their freedom. Sugden's criterion can be used to provide a fully-fledged assessment of some economic situations but remains strangely non-committal towards opportunities for commitment, which implies that nothing can be said about markets for commitment devices—although they conflict with the values of self-liberation that Tyler Cowen extolled. Moreover, Sugden's criterion is limited to the framework of one-person decision problems, whereas it might be interesting to also consider contexts of interaction, as emphasized in the previous section. There is therefore ample room for further conceptual explorations.

0.3 Four conceptualizations of freedom

What we need to go further is a clear assessment of commitments from the perspective of freedom. What is at stake is the possibility to make value judgements about opportunities for commitment from the perspective of freedom, and to understand how and when a loss of control is not necessarily a loss of freedom, as claimed by Sen. Rather than designing a new freedom criterion from scratch, I will consider various conceptualizations of freedom, as I will call them, who already have credentials in the history of political and economic thoughts. I use the term 'conceptualization' to designate the fact that they are neither concepts nor conceptions of freedom, but *general frameworks for evaluating situations in terms of freedom*. I will now explain why I use this terminology.

The distinction between concept and conceptions of freedom is due to MacCallum (1967), who argued, against Berlin's famous distinction between a positive and a negative concept of freedom[77], that the different views of freedom that one can find in the philosophical literature can be interpreted as variations in the specification of the variables of a single concept[78]. The idea is the following (Carter 2022): if philosophers such as Marxists and libertarians really disagreed about the *concept* of freedom—for instance when Marxists claim that poor people are less free, and libertarians deny it—, they would not be talking of the same thing, and their disagreement would not have any political or moral dimension. The fact that their disagreement does have a political or moral dimension shows, *a contrario*, that they share the same concept of freedom, and that their disagreement is about something else, namely the right conception of freedom that one should endorse. According to MacCallum, the term 'freedom' can be analysed as a triadic

[77]'Negative liberty is the absence of obstacles, barriers or constraints. One has negative liberty to the extent that actions are available to one in this negative sense. Positive liberty is the possibility of acting—or the fact of acting—in such a way as to take control of one's life and realize one's fundamental purposes' (Carter 2022). See also Berlin (1969/2002).

[78]The claim that there is only one concept of freedom has been contested in philosophy, especially by using a distinction from Taylor (1979/2006) between *exercise-* and *opportunity*-concept of freedom. The thesis is concerned with an opportunity-concept of freedom, which conceives freedom as having *possibilities* to do or become what one may want. 'If interpreted as an exercise concept, freedom consists not merely in the possibility of doing certain things (...), but in actually doing certain things in certain ways' (Carter 2022). It can be argued that MacCallum's formula does not capture this possible dimension of freedom, which I will not discuss in the thesis anyway.

relation between three different variables. His 'formula of freedom' is the following: 'X is free from Y to do or become Z'. This is the basic concept of freedom that Marxists or libertarians would share. A specification of the nature of the different variables X, Y, and Z can be called a conception of freedom.

- Specifying X means deciding what kind of *agents* are relevant to our conception of freedom. In the thesis, I will only consider conceptions where X are individual beings, and not, for example, groups.

- Specifying Y means deciding what kind of *constraints* are relevant for limiting someone's freedom. We could consider for example that only constraints intentionally imposed by other individuals to restrain us can count as limiting freedom. Hayek (1960/2011) defends such a view. But if a lack of money is also a relevant constraint, as was argued by G.A. Cohen (2011), being poor makes someone unfree, or less free, *as such*—a conclusion that Hayek would not endorse.

- Specifying Z means deciding what kind of *actions* or *outcomes* are relevant achievements for freedom. For example, political philosophers usually consider that only actions should matter for freedom, while Sen considers that 'beings' or outcomes like the absence of malaria are also relevant.

MacCallum's framework is very useful for classifying different conceptions of freedom and comparing them[79]. But it does not directly tell us how we should evaluate situations in terms of freedom, which means producing a *global* value judgement about these situations. The fact that someone is free (or not) *to* go on vacation to the Bahamas, because they do not face the relevant constraints and going on vacation to the Bahamas is a relevant achievement, does not tell us if, under given political or economic arrangements, they are free—or free to a certain degree—, *tout court*. Conceptualizations of freedom would go the extra mile and enable us to pass this latter kind of judgment on a given situation, which allows us to compare and rank a large number of them. The four conceptualization of freedom that I will consider have all already made their way into economics—even if they are not

[79]See Binder (2021b) for an application of this framework to assess markets in terms of freedom.

really in the mainstream—, and are the legacy of solid traditions of political and economic thought. This will be my main justification for using them in the thesis. Following Sen[80], I will not try to make the case that there is a uniquely correct conception or conceptualization of freedom. The fact that these four conceptualizations of freedom have credentials in the history of political and economic thought is a good indicator that they capture important aspects of what people value about freedom, even if much philosophical work has been and can still be done to determine exactly how. And the fact that they deliver different conclusions shows that they are not reducible to one another.

But acknowledging the multiplicity of the dimensions of freedom, and the importance of considering a plurality of conceptualizations of freedom, does not imply that they are identically relevant or valuable. I will borrow from Downding and van Hees three 'criteria for conceptual analysis' (Dowding and van Hees 2007, 148-150) which I will use to evaluate conceptualizations of freedom:

- *semantic criterion.* We would use this criterion if we think that what matters is that a conceptualization of freedom accords with our everyday usage of the term 'free', as we can attest by checking our *linguistic intuitions*. For example, this criterion is used implicitly when we criticize a conceptualization of freedom for implying that someone imprisoned is free, whereas that is not how we would normally use the term. But as Dowding and van Hees note, 'There are so many conflicting intuitions about the nature of freedom that it is often not clear which, if any, of our semantic intuitions should be taken as authoritative'.

- *normative criterion.* This criterion relies on *normative intuitions*. 'The underlying idea is that having freedom is, at least *prima facie*, valuable' and should therefore not have morally repugnant implications. For example, Sen objects to Nozick's doctrine of Lockean rights (which I will describe in the following) that 'Even gigantic famines can result without anyone's libertarian rights (including property rights) being violated' (Sen 1999b, 66), which is a reason to at least revise the con-

[80]'both equality and liberty must be seen as having several dimensions within their spacious contents. We have reason to avoid the adoption of some narrow and unifocal view of equality or liberty, ignoring all other concerns that these broad values demand' (Sen 2009, 317).

tent of Lockean rights, as, arguably, few would maintain that having freedom is a good thing if it is compatible with mass starvation.

- *methodological criterion.* This criterion 'states that there should be a proper fit between the conception of freedom one uses and the research questions being addressed'. It is especially useful in normative economics. A conceptualization of freedom needs to be able to apply to economic situations and be consistent with what economics takes as basic facts or assumptions about agents and their environment. It would also need to be precise enough to enable economists to make recommendations. I will give an example of the use of this criterion in the next paragraph.

There are two possible kinds of judgements about freedom that a conceptualization of freedom can make: either given economic arrangement makes us free or unfree, or they can make us free *to a certain degree*. The latter possibility, which I will call *non-binary*, is familiar in economics: the best economic or political arrangement is the one that maximizes a relevant quantity (discrete or continuous), which reflects the fact that the value considered is best promoted within these arrangements. Utilitarianism is a case in point: the arrangement that maximizes the sum of the utilities of everyone involved is the best possible. That evaluation is then translated into a justification or a reason for action: if this arrangement is in the feasible set of the decision-maker, its implementation is justified, with regard to the value that we consider, and can lead to a recommendation. The second kind of evaluation is *binary*. The goal is to determine whether or not some arrangement verifies a set of properties which gives it a special quality, essential for justification. Take the example of the Pareto rule: a situation is Pareto-optimal if and only if one cannot make everyone better off by moving to another feasible situation. The Pareto rule would state that an economic situation is good if and only if it is Pareto-optimal. This is an instance of a binary evaluation: we can classify every situation as verifying the rule (that is, as Pareto optimal) or not. Just as the Pareto rule, binary evaluation generates a very rough assessment of situations: either the situation considered makes people free, or it does not. This is an example of a methodological criticism: the Pareto rule is not fine-grained enough to deliver a recommendation in most cases, as many different economic situations would qualify as Pareto-optimal. However, the prominence of the Pareto rule in normative economics indicates that binary evaluation can still be useful.

	Non-binary	*Binary*
Commitments incompatible with freedom	Quantitative freedom	Liberty principle
Commitments compatible with freedom	Consumer sovereignty	Lockean rights

Table 2: Four conceptualizations of freedom

A second way of distinguishing between conceptualizations of freedom for my purpose is to ask whether or not opportunities for commitment may count as enhancing individual freedom, or if they undermine it according to this conceptualization. I will now present these four conceptualizations and specify how each of them enables us to answer this crucial question. The two distinctions, taken together, enable us to classify the four conceptualizations as described in Table 2.

This classification, which will be explained in the following, shows that these four conceptualization are significantly different, and provide a different answer to the question of how to regard commitments from the point of view of freedom.

0.3.1 Consumer sovereignty

William Hutt, the first economist to use the term consumer sovereignty, defines it in this way: 'It simply refers to the controlling power exercised by free individuals, in choosing between ends, over the custodians of the community's resources, when the resources by which those ends can be served are scarce' (Hutt 1940, 66). As Desmaray-Tremblay (2020) summarizes, the expression 'connects the liberal value of individual freedom, the commitment to a market society and an appeal to democracy (....) by drawing an analogy between voting in a democracy and buying goods in a market'. In the formal framework of welfare economics, preferences are conceived as ordinal and a preference relation can represent a ranking of bundles of goods just as well as a ranking of candidates, or political proposals, which makes the analogy an identity—in this case, we might as well talk of *individual sovereignty*, as Arrow (2012) did. Consumer sovereignty, as the term is used by economists, refers to the fact that, since preferences are taken as given by economists,

individuals can be seen as completely *free* to rank the alternatives in any possible way they want[81]. And when preferences are conceived as choices that individual *would* make in appropriate circumstances, we have the 'freedom interpretation' of the preference satisfaction criterion, as described by Sugden and McQuillin (2012): if the social planner directly gives me my most preferred bundle of goods, they simply give me the bundle that I would have chosen in my budget set.

Welfare economics' formal framework therefore offers a way to evaluate economic arrangements in terms of freedom. Individuals are free to a degree—the degree to which their preferences are satisfied. But as preferences are reducible to choices, this conceptualization values choices *as such*, not because of what choices express (such as a conception of well-being or an indicator of happiness)[82]. As Sugden and McQuillin note, 'The idea of respecting choices without enquiring into the motivation that lies behind them is libertarian in spirit' (Sugden and McQuillin, 2012, 562). The best arrangement for an individual is the one in which the alternatives that eventually obtain is the one that they would choose, in the largest feasible opportunity set. What matters is only that individuals *eventually* get what they would choose, so this conceptualization of freedom is neutral with processes such as commitments—as long as it leads individuals to get what they would have chosen. The problem, noted by Sugden, is that this conceptualization breaks down when there are inconsistencies in choices. There is no way to identify what individuals would choose in a given set if, for example, the framing of the decision problem is enough to change individuals' decision[83]. Sugden and McQuillin suggest that 'A genuinely libertarian approach (...) would give up the attempt to construct welfare rankings and concern itself only with freedom of choice'. In the case where preferences are unstable or context-independent, an individual might still want to 'be able to satisfy [their] preferences, whatever they turn out to be, as fully as possible' (Sugden 2018a, 97), which having more freedom of choice enables them to do. This would require relying on a different conceptualization, to which I turn now: quantitative freedom.

[81] See Blaug: 'only self-chosen preferences count as individual preferences as yardsticks of individual welfare (in popular parlance: an individual is the best judge of his welfare)' (Blaug 1992, 125)

[82] 'Choice provide appropriate guidance because they are choices, not because they reflect something else'(Bernheim and Rangel 2009, 52)

[83] See Kahneman (2013, 363-374).

0.3.2 Quantitative freedom

Each time someone says that they would be *freer* in one place than in another, or that a society, marketplace, etc. offers more freedom than another, they refer to freedom as a quantity that can be maximized. As Carter, a proponent of this conceptualization of freedom, explains, it is 'the view that it is important to be able to say *how free* an individual or society is—sometimes absolutely, sometimes comparatively, but nearly always in an 'overall' or 'on-balance' sense' (Carter 1999, 3). What matters is not *if* we are free to do particular things, like saying our mind, or voting, but *how* free we are in this overall sense that would enable us to formulate a global evaluation of economic situations. Freedom is something good (at least *prima facie*), and it is always better to have more options, which conflicts with commitments. As Carter recognizes, this conceptualization is not very popular in political philosophy[84], probably because of widespread scepticism[85] about the possibility of measuring it in a convincing way[86]. But the scepticism regarding the measurement of quantitative freedom among academic philosophers coexist with extensive political discussions about the extent of freedom that we can enjoy in some societies compared to others. The literature on freedom of choice also refers to freedom as a measurable quantity, but despite its achievements, it did not produce a consensual framework for evaluating arrangements in terms of (quantitative) freedom, which could serve as an alternative to cost-benefit analysis or measurement of consumer surplus in welfare economics.

There are many reasons to be sceptical about the possibility of providing convincing measures of quantitative freedom—even if we set aside normative objections concerning the importance or the true nature of freedom.

- First, as Sen remarks, it is much easier for an economist to observe actual choices than to observe opportunity sets. Observing opportunity sets is much more requiring in terms of information[87].

[84]Dworkin (1985), Kymlicka (2002), Hart (1973), O'Neill (1980), among others, have rejected it.

[85]See in particular O'Neill (1980) for a detailed argument.

[86]See Carter (1999) for an extended discussion on this question.

[87]'The capability set is not directly observable, and has to be constructed on the basis of presumptions (just as the 'budget set' in consumer analysis is also so constructed on the basis of data regarding income, prices and the presumed possibilities of exchanges). Thus, in practice, one might have to settle often enough for relating well-being to the

- There are also conceptual problems to be solved. When individuals' choices are interdependent, the availability of the alternatives in someone's opportunity set would depend on the choices that others make in their opportunity set—otherwise, some potential choices could be mutually incompatible—but there is no obvious way to take these interdependencies into account[88]. Another conceptual problem is related to the need to incorporate preferences defined over alternatives in the evaluation of individuals' opportunity sets, as mentioned in section 1. This raises two problems.

  - The first is that preferences may be unstable, or dependent on the context. In particular, if preferences are adaptive—which means that alternatives are deemed less valuable when they are not in the opportunity set (as evoked by the expression 'sour grapes'), especially for people who are particularly deprived and have few opportunities—then we would systematically underestimate the gain in freedom that the addition of an alternative can make to a deprived person's opportunity set, compared to someone who is not deprived[89].
  - The second problem, emphasized by Binder[90], is that it makes it conceptually difficult to identify cases of paternalism. One of the basic conditions for paternalism is that such acts interfere in some way with someone's freedom or autonomy[91]. But suppose that our account of freedom implies that an action that someone

achieved—and observed—functionings, rather than trying to bring in the capability set' (Sen 1995, 52).

[88]Pattanaik and Xu (2018) explore multiple ways to delineate opportunity sets when choices are interdependent, but their paper remains inconclusive. There are a few proposals for measuring individual freedom in strategic interactions representable as *games*, where the outcome that one gets depends on the actions chosen by others: see in particular Bervoets (2007), Ahlert (2010) and Sher (2018). However, these papers do not attempt to connect the proposed measures to particular philosophical conceptions. For a proposal that defines a 'freedom function' capturing negative freedom in *simple games*, see Braham (2008).

[89]Elster (1983) brought the theme of adaptive preferences to light. For more on this issue, see Costella (2023).

[90]See Binder (2015, 31-33) and Binder (2021a).

[91]According to Dworkin's classic definition of paternalism, there is paternalism when someone (1) interferes with the freedom (or autonomy) of someone else, (2) without their consent and (3) to promote their good (Dworkin 1972).

> could have performed (Binder's example is 'wearing a headscarf') is taken to be worthless because no one—or no one deemed reasonable—values it: interfering with this action would thus not significantly alter this person's freedom and would not count as paternalism. Certain acts of paternalism are 'defined away' if the value of alternatives should count when evaluating freedom.

Recently, Sugden (2018a) has revived the attempt to use quantitative freedom to evaluate economic arrangements. Sugden's ambition is to rewrite the theorems of welfare economics by using an opportunity criterion (instead of the traditional preference satisfaction criterion), which bypasses the problem of unstable preferences and context-dependence. The conceptual problem of interdependence is avoided by defining a fairly complex 'interactive opportunity criterion' applicable to market transactions, which seems to make sense at the level of the group[92], rather than that of the individual:

> In every competitive equilibrium, individuals' opportunity sets satisfy a condition that I will later define more formally as the 'Strong Interactive Opportunity Criterion' (...). This condition requires that every group of individuals has the collective opportunity to make any feasible transaction among themselves which, given the assumed desirability of money, they might find mutually acceptable. (Sugden 2018a, 111)

The problem that the actual opportunity set faced by individuals might be unobservable is avoided by Sugden, who bluntly asserts that 'opportunity is an open-ended concept: often, we cannot specify in concrete terms what a person does or does not have the opportunity to do' (Sugden 2010, 48). Under a conception of 'opportunity as mutual advantage', opportunity is not a quantity that can be *measured* but it can still be *defined*, so as to show the ability of market outcomes to satisfy the interactive opportunity criterion[93]. At bottom, when we ignore all the complexities brought about by interdependencies or dynamic decision problems, Sugden's criterion seems to

[92] 'People *collectively* can expect the market to provide them with a rich array of opportunities to transact with one another on terms that they might find *mutually* acceptable' Sugden (2018a, 109).

[93] 'We cannot say whether one's person opportunities are greater or less than another's. However, there is a sense in which we can say whether, within a given economy, all feasible opportunities have been made available' (Sugden 2010,55).

be based on an inclusion rule[94], which states that we have more freedom when we have more alternatives to choose from, independently of the preferences of the chooser[95]. This avoids the problem of paternalism identification, as the criterion is purely neutral and counts *any* alternative as contributing to the freedom offered by the opportunity set in the same way. The challenge faced by Sugden seems to be that either 'opportunity as mutual advantage' reduces to the intuitive idea that it is good to have more alternatives to choose from (as defined by the simple inclusion rule), but then it cannot easily be applied to most economic situations (because of interdependencies), or it is different and more complex, and then the difficulty is to show why it is particularly desirable.

0.3.3 Mill's liberty principle

The way economic or political actors talk about freedom is sometimes quantitative, as I remarked earlier: people are freer in some country rather than another, etc. But it can also be simply, as I called it, *binary*: someone is a free man or woman and lives in a free country, located in the 'free world', trade goods on a 'free market', absolutely. A common way to make sense of these expressions is the observation that some rules may be satisfied or not, which grants some political or economic arrangements—and the people acting under them—the status of being free. I will explore in the next two subsections two different ways of specifying these rules that allow arrangements satisfying them to qualify as free:

- *basic rights.* The first is to specify a set of rights or basic liberties that people *should* have and check whether *all* of them are respected in a given situation. If this is not the case, then people are not free under this arrangement. I will consider in the next subsection Nozick's doctrine of Lockean rights, which exemplifies this approach in a particularly extreme way since every right of every individual has to be respected for a society to count as free. The difficulty is to provide an adequate normative justification for the set of rights and basic liberties whose respect is necessary and sufficient to make individuals free. I will present Nozick's doctrine of Lockean rights in the next subsection.

[94] See Sugden (2018a, 84-85).

[95] In other words, an opportunity set gives more freedom than another if the latter is included in the former.

- *harm principle.* The second, which is less common, is negative. It focuses on identifying *actions* that threaten freedom. Once these actions are identified, one can say that people are free under some arrangement if individuals' freedom is protected against these actions. This second approach is based normatively on the idea of a presumption of liberty: as Mill puts it, 'in practical matters, the burthen of proof is supposed to be with those who are against liberty' (Mill 1869/2006, 134). Individuals should be left free to do what they want, provided that they do not commit actions that threaten others' freedom. In Mill's version of this approach, which I will present now in more detail, the harm principle—sometimes also called the 'liberty principle'—states that there is a justification to restrict someone's freedom if and only if their conduct can be said to *harm* others. The observation of this liberty principle, which is meant to *protect* the exercise of freedom, would make a society free.

Some passages from Mill's *On Liberty* may suggest, however, that he actually supports the first approach:

> there is a sphere of action in which society, as distinguished from the individual, has if any, only an indirect interest; comprehending all that option of a person's life and conduct which affects only himself, or if it affects others, only with their free, voluntary and undeceived consent and participation. (...) This, then, is the appropriate region of human liberty. (Mill 1859/2006, 18)

Mill proceeds by enumerating three categories of liberties that this region of human liberty comprises: the 'inward domain of consciousness', the 'liberty of tastes and pursuits', and the 'freedom to unite' with other people. 'No society, in which these liberties are not, on the whole, respected, is free' (Mill 2006, 19). According to this line of thought, conduct can be either self-regarding or other-regarding. Society would be free if and only if the private sphere of self-regarding actions is kept unobstructed. This idea has motivated Sen's formulation of rights in his paper about the impossibility of a Paretian liberal. But this interpretation raises several problems. The first, which was often noted, is that it is very difficult to define a sphere of self-regarding conduct, as nearly every action that we can take impacts others in some way. This would make the 'region of human liberty' very small, and the 'liberty principle' is not up to the task of representing or embodying

Restricted conduct	*Region of human liberty*	
Other-regarding conduct		Self-regarding conduct
Harm	Offense	

Table 3: Liberty principle and self-regarding conduct

our concern for freedom. The second problem is that something needs also to be said about freedom in the sphere of other-regarding conduct: mere *offence*, according to Mill, in contrast to harm, is not something that society should be prohibiting or sanctioning. While conduct causing offence is not self-regarding, it is something that would belong to the 'region of human liberty', however unpleasant it is.

A better way to interpret Mill would thus use the liberty principle—which separates actions harmful to others from actions which are never to be restricted— to define negatively the 'region of human liberty'. As emphasized by Lovett, the definition of this region is 'a by-product of the argument, not its basis', because 'conduct within the [private] sphere is exempt from social regulation because it is harmless, not because it is (...) private' (Lovett 2008, 126). Indeed, one can find passages from Mill that support the idea that in a technical sense[96], conduct is self-regarding only when it is *not* harmful to others. But since this is not really compatible with ordinary ways of thinking, David Brink suggests keeping the two distinctions independent: conduct is either self-regarding or other-regarding, *and*, other-regarding conduct may cause harm or not. If some conduct cause harm, that makes it permissible for society to restrict it, while it is impermissible to restrict all other kinds of conduct. What matters here is that the sphere of action which is of interest for evaluation in terms of freedom is delimited by the application of the harm principle, and not by the assumption that some kind of conduct is self-regarding. These nestings are represented in Table 3.

In essence, the ability to conduct evaluation in terms of freedom hinges on the feasibility of defining and identifying instances of harm. An important qualification made by Mill is the principle that *volenti non fit injuria*, which says that what is collectively agreed upon by individuals does not constitute harm, and should not be interfered with by society. The liberty principle thus covers freedom of contract, by which someone can agree to restrict his own freedom to get something in return from others. This raises the question

[96]On this, see Brink (2013, 140).

of whether the harm principle also covers commitments. In the case of the liberty principle, we may think that since acts of commitments appear to be self-regarding, and in any case do not harm others, they are protected by the liberty principle. But what Mill says about slavery contracts suggests the contrary, as I explain in chapter 5. The argument made by Mill to exclude slavery contracts entails that not only these contracts but *any* contract which would imply that commitments may be enforced (for example by preventing someone from renouncing what they had vowed to do) should be considered null and void. As David Archard (1990) has shown, this is not, as some have said, an artificial, *ad hoc* exception to the liberty principle, but a consequence of the fact that this principle is *meant* to protect freedom. This purpose would be contradicted by allowing individuals to lose the very freedom the principle is meant to protect.

0.3.4 Lockean rights

The (negative) conception of freedom implicit in Mill's liberty principle is that individuals are free when they are *left* free to do what they want—whether or not that they *can* do it—as long as it does not harm others. This does not make it necessary to define opportunity sets or identify preferences—one only has to identify instances of harm. Another way to delineate the sphere of what individuals *should* be left to do to be free is to define a set of rights that should be respected. Within the sphere delimited by these rights, individuals are free to act as they please. Giving someone such a right, as Nozick describes them, imposes some constraints over the actions of every other individual: 'the right of others determine the constraints upon your actions' (Nozick 1974, 2), hence the name *rights as side-constraints*. Once every such right has been specified, the set of permissible actions is determined for every individual, among which they are free to choose[97]. This formulation implies that no trade-off can ever be made between respecting the rights of someone and and those of another person—this excludes that

[97]Ideally, as mentioned in the first section of the introduction, the exercises of rights fix all relevant 'features of the world' and would leave nothing to be deliberated or chosen collectively. The claim that such a 'right structure' will settle all issues that individuals may have in society and leave nothing to public deliberation or social choice in a general sense, is impossible to reconcile with other conditions such as Pareto or symmetry in the assignments of individual rights. For such an impossibility result, see Braham and van Hees (2014). This constitutes a significant limitation of Nozick's approach.

we could, for example, violate someone's right to prevent *multiple* violations of other people's rights[98]. This impossibility, which puts this conceptualization of freedom at odds with quantitative freedom[99], reflect the libertarian view that no one should have their freedom sacrificed—even in the slightest way—for the sake of others.

These rights do not act only as constraints on the actions of other individuals: they are also, in Nozick's framework, constraints on possible interventions of public authorities. No intervention that entails the violation of at least one individual right is justifiable. This latter feature comes from the fact that these rights are *Lockean rights*—rights that play the role that Locke has assigned to them—, which means that:

- individuals rights in general, and property rights in particular, are conceived as moral or *natural*, which means that 'they find their justification and authority prior to their recognition by political or legal actor'. These rights are pre-political and exist independently of the political realm.

- public authorities derive their legitimacy from their role in protecting these rights and 'may therefore not violate individuals rights without losing their legitimacy' (Levy 2017, 22).

The fact that these Lockean rights are respected implies that society is, at the same time, *just*—because rights are respected—and *free*—because these rights enable individuals to enjoy freedom, as Locke defines it: 'a liberty to dispose, and order as he lists, his person, actions, possessions, and his whole property, within the allowance of those laws under which he is' (Locke 1713/ 2016, 29). At some point in time, Lockean rights specify what someone's possessions are and what actions he can do. This involves also a description of how these legitimate possessions and permissible actions can *change* through time—whether, how, and when individuals may transfer their possessions, and commit themselves to do certain things. The content of Lockean rights

[98]This would be the implication of an 'utilitarianism of right' that Nozick rejects (Nozick 1974, 28).

[99]Quantitative freedom may allow, if interpersonal comparisons of freedom were possible, that someone's freedom is decreased provided that the increase of others' freedom is sufficient to offset this loss. Thus the classical objections to utilitarianism (that it does not treat individuals as separate persons, as Rawls 1971 insisted) would also apply to quantitative freedom.

therefore specifies the form under which exchanges and contracts are permissible. The exact nature of Lockean rights would depend on one's moral theory about what we have a moral right to do[100]. This offers some flexibility, which can be exploited to specify a conception of freedom that can address the challenges of evaluation in normative economics[101].

Under a narrow libertarian interpretation, Lockean rights are full *property rights* over oneself and things that one has justly acquired. They can always be exchanged and transferred at will, provided that individuals consent to it, and no one can ever interfere with someone's use of their property. It is therefore permissible for anyone to dispose of themselves, or bind themselves, in any way they want, to an extreme degree, which allows for assisted suicide or slavery contracts, as Nozick recognized. But this is not necessarily the only form that Lockean rights can take, even if they only caught the interests of libertarians. In any case, to design a fully-fledged economic doctrine, one needs to explain, as Locke and Nozick did, how possessions can be justly acquired, and which actions someone can be allowed to perform—which is a difficult task. But if our purpose is only evaluative, as is the case here, we need not concern ourselves with this problem. An interesting property of Lockean rights is that permissible commitments and transfers, when they are performed, *change* the set of possessions that each person has and the set of actions that they may perform while *preserving* the justice of the society that enforces these rights. As justice is preserved, freedom is also preserved, because of the tight connection between the two in a Lockean framework. The result is that, whatever may be the *degree* of freedom that we have achieved at some point in time in our society, we can be sure that it is preserved through time if individuals only do what they have the right to do. This enables us to make judgements about freedom even if we are not able to observe or delineate opportunity sets. Chapter 2 explores the consequences of this property for allowing a public and coercive intervention aimed at producing public goods.

The first two conceptualizations of freedom—consumer sovereignty and quantitative freedom—have already been extensively developed in normative

[100]Locke's conception of freedom belongs to the category of what is called 'moralized conceptions of liberty', which 'builds morality into the very concept of liberty, thereby ensuring that it is intrinsically normatively significant and that it can, consequently, play a central justificatory role in moral and political theorizing' (Bader 2018, 2).

[101]Chapter 2 makes some steps in this direction.

economics. I have discussed some of their limitations. The latter two, despite having generated a huge literature in philosophy, have not received similar consideration in economics. These conceptualizations rely on a 'negative' and 'moralized' approach focused on identifying what individuals should be *left* to do, which makes the problem of identifying the alternatives in complex economic situations where individuals are interdependent disappear. Their downside is that evaluation can only be binary. It would thus be necessary to supplement an evaluation in terms of freedom with other evaluative tools that could generate finer rankings of situations. Chapters 2 and 5 will show that significant normative conclusions can still be drawn from these conceptualizations, which have implications for public debates about the economy. The next section will describe the approach that I will take to contribute to these public debates and to analyse the role that values can play in normative economics.

0.4 Methods, approaches, results

This thesis belongs to the field of philosophy of economics. Philosophers and economists sometimes work on the same topics, as this introduction has already made clear, and sometimes with the same methods[102]. Even though economists do not always acknowledge it[103], there is a lot of overlap between the two. Philosophy of economics is not just an extension of the field of philosophy of science (concerned with analysing the methods used by the various sciences) to economics, since it also covers many areas where economics and other fields of philosophy, such as philosophy of action, ethics and metaphysics, intersect. As a sub-discipline distinct from economic methodology, it is relatively new[104]. As Hands (2018) explains, past economic methodologists either did not mention philosophers (as exemplified by Friedman 1953) or 'tried to apply the arguments of particular philosophers of natural science directly to economics' (as exemplified by Blaug 1980/1992). However,

[102]Especially in the area that has been called 'formal ethics', which is concerned with the application of formal, mathematical tools to ethics

[103]Robins' influential essay, for example, takes great care to separate economics from moral philosophy (Robbins 1932/1984)

[104]Economists never ceased to discuss methodology but as Hausman (2021, 15) notes, 'In the 1970s, there was not yet a great deal of discussion across the boundaries between economics and ethics'. It is only in the 1980s and especially 1990s that philosophy of economics emerged as a sub-discipline, according to Hausman.

> Since the mid-1980s there has been a renaissance in the interaction between economics and philosophy. The traditional approach to economic methodology continues to produce viable research, but economics and philosophy are also interacting in many other, new and important ways. Philosophy of natural science is no longer the only relevant set of philosophical ideas—ethics and ontology have both returned to the scene—and the intellectual dynamic is now one of bilateral exchange rather than economists simply borrowing ideas from one corner of the philosophical shelf. (Hands 2008, 411).

The thesis contributes to this bilateral exchange in two different respects: by following in the footsteps of philosophers of economics interested in describing the role of values in economic theories and practices and by constructing and analysing arguments establishing the legitimacy and value of public interventions. I will present in the following the two literatures that provide the research framework of this thesis: the methodological literature on the role of values in economics, and the literature on 'economics and ethics' or 'economic ethics'. Finally, I will present the chapters of the thesis and their conclusions. But before that, I will briefly explain how the discussion of the questions I mentioned is conducted and how results are obtained—which is by use of *arguments*.

Philosophers are trained to construct and analyse *valid* arguments, whose conclusion is necessary entailed by their premises, especially when formulated in informal language—which is the language of the individual or collective decision-makers who (hopefully) read and use what economists have to say about the world. In the following, I will thus defend the idea that the task at hand requires an interdisciplinary perspective. As Atkinson (2001), Hausman et al. (2016), and Alexandrova (2016) have argued, the division of labour between economists and philosophers that Robbins (1932), among others, envisioned—economists being concerned with the *means*, philosophers with the *ends* of human action, and therefore with values—is neither feasible nor desirable. This means that the task I will describe in the following requires us to be both economists and philosophers to be properly performed.

0.4.1 Analysing arguments

All chapters of the thesis rely on the analysis of arguments.

- Chapter 1 examines what I called the simple argument for markets, which makes the case that opening a new market provides more freedom to act for everyone.
- Chapter 2 revises the public good argument to address the concern that public provision of public goods may involve coercion.
- Chapter 3 scrutinizes the anti-paternalist argument which concludes that paternalistic interventions are never optimal.
- Chapter 4 builds an argument to show the incompatibilities between anti-paternalism, a freedom criterion for evaluation and preferences for commitment.
- Chapter 5 constructs and analyses possible arguments for regulating markets for commitment devices.

Each of these arguments has normative premises and descriptive premises. The point of the exercise is to examine these arguments as a whole, by relying on philosophy for the normative perspective that political philosophy or ethics gives to evaluate situations, and on economics for the models it provides to describe these situations—which permits to *apply* (through substantial adaptation to the model) the normative perspective. While (simple) economic models are considered in each chapter, the discussion is conducted in informal language. The main reason why this is the case is that public debates rely on arguments expressed in informal language because, as many philosophers of science have pointed out[105], concepts expressed in informal language provide *reasons* for action or political intervention, which is ultimately what normative economics call for. Economic models are often implicitly embedded into an argument structure (as in the case of the public good argument, which I will describe later in the section), which is often only sketched in economics papers but need to be made explicit to give reasons for interventions[106].

[105]Dupré makes the case that even when social sciences can avoid using the value-laden terms of the ordinary language, relying instead on technical terms, it would need to get back to it at some point to connect its discourse to 'human concerns': 'there are plenty of more or less wholly value-free statements, but they achieve that status by restricting themselves to things that are of merely academic interests to us.' (Dupré 2017, 40).

[106]Section 2 of Chapter 2 of the thesis consider the special case of normative models, which are meant to be used to derive explicit normative conclusions. See Beck and Jahn (2021) for an analysis.

An argument is called 'sound' if it is 'valid' and if its premises are at least plausible. Validity is a logical condition: an argument is valid if (and only if) its conclusion cannot be false if its premises are true. In the particular context of the arguments I describe in the thesis, the plausibility of the premises depends on the appeal of normative principles that enter the premises and the fit between the economic model and the particular situation considered. This defines three different tasks:

- checking the logic (or validity) of the argument;
- discussing the appeal of the normative principles;
- discussing the fit of the model with reality.

Only the first two tasks are undertaken in the thesis, as it relies on philosophical expertise.

Tests of validity matter because they enable to identify *missing premises*, which are overlooked by those who made the argument. Missing premises are crucial in the analysis of arguments. Once properly constructed and added, they make the argument valid, but they also open the way to formulate innovative criticism because the argument may no longer appear sound if the reconstructed premises are not plausible. This task, which may appear only negative, is the occasion to discover new possibilities, if criticism can morph into a positive alternative proposal. That is the path I take in chapter 3, exploring a new kind of paternalism which I call 'strategic paternalism'. Another virtue of constructing valid arguments is to make explicit incompatibilities, to map out the logical space of premises and principles which are compatible together. This is the path taken in chapter 4, where I show in particular that anti-paternalism is not compatible with allowing for a stringent preference for commitment.

Discussing the appeal or plausibility of normative principles is a less straightforward task. I have described at the beginning of the third section of this introduction three criteria that help evaluate such principles: the semantic, normative and methodological criteria. Applying these criteria can serve to filter out principles which, respectively, clash with linguistic intuitions, with considered moral judgements or are simply irrelevant to the task ahead. Principles borrowed from philosophical or ethical schools of thought that have withstood these three kinds of criticism can be deemed, at least provisionally, to have some normative appeal. If only these principles enter

into the normative premises of a valid argument, we have produced a sound argument that justifies the conclusion we want to assert—provided that the descriptive premises are also plausible. The difficulty of the philosophical task is also to make these arguments non-trivial, by which I mean that the argument derives a seemingly strong (or surprising) conclusion from weak (or obvious) premises. The larger this gap between the conclusion and the premises, the better the argument is.

0.4.2 The role of values in normative economics

An important aspect of social sciences like economics is that they enable people to learn useful lessons about what matters to them in their social life. Someone's wealth or income is obviously something of interest to him, which might be enough to stir curiosity about economics. Employment, income and GDP are everywhere in the public debate about the economy, as well as in economics. But since it is not GDP as such that interests people, but what they collectively can *do* with it, and how better off they are with it, welfare economics appears as a natural extension of economists' theories about wealth or income. As we have seen, economists have traditionally adopted a minimalist, non-committal conception of welfare when it comes to evaluating how better off people are. They are better off insofar as they get what they prefer and would—in appropriate circumstances—choose. The fact that many economists have agreed on a methodology that implies to deferring to individuals' preferences, rather than selecting a particular conception of the good of their choice to evaluate how individuals fare is indicative of a commitment to anti-paternalism. There are many possible reasons behind such anti-paternalism. It can be motivated by a concern for neutrality or autonomy. One reason to defer to individuals' preferences is that it enables economists to minimize the number of value judgments that they have to make to evaluate someone's situation (which is what Haybron and Alexandrova 2013 have called 'normative minimalism'). This enables economists to be neutral with respect to individuals' conception of the good[107]. A different reason to be anti-paternalist would rely on a concern for individuals' autonomy: individuals *should* be free to satisfy whatever preferences they might want to have. But this is a value commitment that economists usually avoid making.

[107]See chapter 4 for more details.

Independently of the particular causes which have historically led economists to endorse normative minimalism[108], there are good reasons to embrace a value-free ideal in science, as much as it is possible. A value-free ideal maintains that scientific theories should avoid, as far as possible, containing or implying value judgements. Value judgements can be defined as 'decisions that involve the weighing of values' (Eliott 2022, 6). Philosophers of science use the term to refer to any choice that involves such a 'weighing' between different desiderata, which can relate to what is sometimes called 'epistemic values': fertility, coherence, explanatory power, etc., as well as to 'non-epistemic values', such as public health, environmental conservation, social justice, etc. Economists usually use the term only in relation to these non-epistemic values, and that is also what proponents of the value-free ideal are concerned with. This introduction adopts this understanding of the term 'value judgement'. Eliott (2022) enumerates three reasons for supporting the value-free ideal:

- *wishful thinking.* The reference to values may detract scientists from the goal of pursuing truth—and instead permit them to serve ideological commitments by accepting hypotheses that promote non-epistemic values rather than those that seem true, that have more explanatory power, etc. This would lead to wishful thinking or even to scientific fraud. The case of Lysenkoism is often mentioned to point to the dangerous implications of the claim that science is necessarily politically-oriented, or class-oriented.

- *autonomy of the decision-maker.* A value-free science would also protect the autonomy of decision-makers. They may not share the same values as the scientists. If a decision-maker—individuals or public authorities—makes their decision based on a theory which incorporates some hidden value judgement, it may alter this decision in a sense which is not reflective of the decision-maker's own values. Think, for example, of formulating predictions that only consider worst-case scenarios. The decision-maker is therefore at risk of being overcautious, compared to what would be their own assessment of the risk-benefit trade-off.

- *public trust.* Keeping values out of science may preserve public trust: the image of a disinterested scientist would be essential to establish scientific discourse as trustworthy. In incorporating non-epistemic values

[108]See Baujard (2017) for a detailed history.

in their reasoning, scientists might undermine science's trustworthiness in the eyes of the public. The argument is empirical, and according to Elliott, not substantiated even if it is plausible.

With regard to these two latter reasons, what the argument should conclude is not exactly that no value judgement should be made, but rather that they should not be kept *hidden*, which would induce in errors decision-makers or individuals who expect factual accounts on certain matters. It is now consensual among philosophers of science that science—and economics in particular—cannot really be value-free. But scientists can attempt to be as transparent and precise as possible about these inevitable value judgements, to protect the autonomy of the decision maker, or preserve public trust as far as possible.

What philosophers of science (and in particular, philosophers of economics) have emphasized is that this specific task of making value judgements explicit and relevant cannot really be organized along the lines of a simple division of labour, which would make scientists in charge of determining *where* a value judgement is needed and philosophers, or decision-makers, or the general public, in charge of deciding *which* value judgement should be adopted. Alexandrova, for instance, objects to this view that 'It ignores or devalues scientists' knowledge about values, which they have acquired in virtue of their knowledge of facts. This knowledge enables them to make better normative choices *qua* scientists' (Alexandrova 2016, 11). Knowing facts and scientific methods helps make better, or more appropriate, value judgements, in part because it helps to know what is at stake. This view calls for an interdisciplinary perspective which is adopted in this thesis: both economics and philosophy are needed to assess what is at stake when making value judgements. Deciding to make such a value judgement is based on an argument: from a philosophical perspective, a value judgement should not be an arbitrary pronouncement but a conclusion obtained from premises that stem from methodological requirements, as well as a normative commitment (as will be exemplified in the thesis).

Until recently, works in philosophy of economics about the role of values were limited to adding to the debate about the status of economics as a value-free science and the division of economics into a positive and a normative branch. But in the last decades, several attempts have been made to analyse the value-ladenness of economics research beyond these two traditional top-

ics[109], especially in relation to evaluation, as exemplified by the work of Anna Alexandrova. One of the goals of this thesis is to prolong these attempts. Values do not enter normative economics only through the choice of a normative criterion (like preference satisfaction, or freedom), but also through the attitudes of normative economists towards individuals' own values and preferences (through anti-paternalism), and in the process of choosing and conceiving normative models of what matters to individuals. The thesis will make three contributions to clarifying the role of values in normative economics by analysing (1) the relationship between the representation of the economic agent and the attitudes of economists towards paternalism, (2) the relationship between anti-paternalism and the adoption of a normative criterion for welfare economics, (3) the value-laden character of normative models of freedom or opportunities. A summary of the contributions is given in the last subsection, which presents the chapters of the thesis.

0.4.3 Ethics and economics

The thesis also makes contributions to the line of research that lies at the intersection of ethics and economics. There exist two handbooks devoted to the topic (Pail and van Staveren 2009, White 2019), but the editors seem reluctant to call 'ethics and economics' a field. According to White (2019), two distinct approaches can be identified. The first, and most visible, follow Amartya Sen and his criticism of the 'narrrow focus on the maximization of utility of welfare to the neglect of concepts such as virtue, rights, dignity or justice'. The thesis takes up this critical approach and explores its consequences. Another approach 'consists of the wealth of research into ethical behaviour on the part of mainstream economists', such as research on altruism and social preferences, which is also touched upon in the thesis. The intersection between economics and ethics is far from empty. But 'ethics and economics' is a mere juxtaposition. Binder and Robeyns (2019) use the expression 'economic ethics' to refer to a special region of applied ethics[110], which would exist alongside medical ethics, bioethics, or ethics of technology. But the expression is not extensively used and its exact relationship with normative economics is not clear, since many contributions that lie at

[109] See Malecka (2021) for references.

[110] Beauchamp (2003, 3) gives this definition of applied ethics: 'philosophical methods to treat moral problems, practices, and policies in the professions, technologies, government, and the like'.

the intersection of ethics and economics are not really 'applied' but concern what is sometimes called the 'ethical foundations of economics'. Hausman et al.'s (2017) widely used book does not contain any reference to the term. Wight (2015) entitled 'Ethics in Economics' his introductory book.

The lack of precise terminology is mirrored in the lack of a clear methodology[111]. Hausman et al. devote a lot of effort to making the case that economics (positive or normative) would benefit from recognizing and adopting an ethical perspective. But they never really indicate precisely *how* ethics should be done by or in relation to economics. They note in passing that 'One addresses moral questions (...) by making arguments' (Hausman et al. 2017, 9). I will try in the following to elaborate more on this indication by way of an example: I will consider the 'public good argument' (discussed in chapter 2) that economists sometimes make to justify the public provision of public goods, which addresses an important political question concerning the role of the state while leaving major ethical concerns aside.

The public good argument has a long history in economics. It has been formulated by Adam Smith and John Stuart Mill, among many others. For Mill,

> It is a proper office of government to build and maintain lighthouses (...) for since it is impossible that the ships at sea which are benefited by a lighthouse, should be made to pay a toll on the occasion of its use, no one would build lighthouses form motives of personal interest, unless indemnified and rewarded from a compulsory levy made by the state (Mill 1848/1909, 976).

Samuelson, who defined the modern concept of public good[112], notes that the existence of public goods entails that '*laissez-faire* can not be counted on to lead to an optimum'. He adds that 'there is a *prima facie* case (...) for social concern and scrutiny of the outcome; but that does not imply outright state ownership in every case public regulation. The exact form in which the social concern ought to manifest itself depends on a host of considerations that have to be added to the model' (Samuelson 1972, 52). The structure of the public argument can be analysed (see chapter 2) as depending on two descriptive or 'positive' premises: a premise that identifies what can be called

[111] See, however, Fleurbaey (1996) on the division of labour between economics and philosophy regarding the axiomatic methods to define and characterize criteria of justice.

[112] Richard Musgrave also contributed to the definition. On this, see Desmarais-Tremblay (2017).

a 'public good problem', namely, in Mill's example and formulation, the fact that building a lighthouse is not in the best (personal) interests of ship owners even though it is needed; and a premise that makes the case that any other solution than a form of 'compulsory levy' is unavailable. Economic models, as well as empirical studies, can make clear whether these two descriptive premises really hold[113]. However, economists concerned with the provision of public goods do not usually discuss a third normative premise, which is that it is better, on the whole, that the public good is provided, even if the use of coercion is necessary to attain this result. This third normative premise can only have some appeal if citizens or decision-makers are willing to approve the trade-off that it implies: less control against more welfare.

Models of public good provision are of interest in the public debate about public goods to the extent that they are useful to make or rebut the public good argument, which concludes that the state should intervene to force individuals to contribute to producing public goods. Because of the (often implicit) normative premise of the argument, it may be that people or decision-makers would disagree with the public good argument while agreeing on the other (descriptive) premises, that is, basic facts about the nature of the public good and the feasibility of public good provision. People, might in particular, not be convinced by an argument that does not address the reluctance that one may have to force citizens to contribute to the financing or production of the public good, even if it is the only way to attain an optimal provision[114]. As was argued earlier, a strict division of labour between economists (concerned by the descriptive premises) and philosophers (concerned by the normative premises) is not likely to be very productive if their set of skills and interests are mutually exclusive[115]. Examining the *whole* of the public argument thus requires an interdisciplinary approach, as the

[113]More precisely, a model is needed to identify situations as raising a public good problem, and used to make the case that no mechanism can serve to provide the public good by voluntary, private provision.

[114]This is especially the case if the addressee of economist discourse is not the more or less fictitious figure of the social planner but actual citizens.

[115]Atkinson gives two reasons why a division of labour between economists and moral philosophers is not desirable: (1) 'many of the key issues can only be understood in the context of relatively sophisticated economic models.', and (2) 'the relation between economics and ethical principles is not linear but rather iterative. Examination of the implications of moral principles in particular models may lead to their revision. By applying ethical criteria to concrete economic models, we learn about their consequences, and this may change our views about their attractiveness' (Akinson 2001, 204).

ethical criterion or principle involved in the normative premise of the argument has to be tailored to fit the model of public goods, to which it is applied.

Arguments matter for economics. Economic methodology puts a lot of emphasis on models, but models can only serve their purpose of guiding public action if their results and assumptions are embedded in an argument that can eventually be formulated in informal language, because a technical language cannot by itself provide reasons for action. This is the 'economics' part. 'Ethics', on the other hand, is concerned with the normative underpinnings of the argument. An argument is not likely to be very convincing if it appeals only to a narrow set of values and neglects every other concern that we may have. But at the same time, the appeal to values that normative economics usually neglect, such as freedom, rights, dignity, etc. is empty talk if it is disconnected from the economic reality—as described by economics models. Informal arguments are not substitutes for models. Economics and ethics cannot therefore be juxtaposed, and a major task that this thesis takes up is to find out how various conceptualizations of freedom can apply to the models to make a convincing argument.

0.4.4 Presentation of the chapters

Chapter 1. A problem that normative economics has to confront is the necessity of making value judgments when designing normative models. Normative models can be defined as 'the class of formal models aimed at providing normative guidance' (Beck and Jahn 2021). The use of value judgments in the design and application of these models has been relatively unexplored in the philosophy of science. An interesting exception is Sugden (2003), who makes the point that it is impossible to define a so-called 'pure quantity' model of individual opportunities that would be entirely neutral concerning the identification and evaluation of opportunities. A pure quantity model is supposed to reflect the opportunities available in the world as they are and avoid any value judgment. Sugden suggests that this is an impossible task. I show, in a concrete case, why this is so: based on what I call the 'simple argument for markets' (particularly advocated by Arrow), the opening of new markets provides people with more opportunities, regardless of their preferences. With a new market, they can now buy and sell where they could not trade before, or at least not for money. This judgment should pose no problem from the perspective of a pure quantity model of opportunities. However, if we revisit

the debates on the commodification of blood transfusions that pitted Arrow against Titmuss, we see that they hold incompatible representations of the opportunities available to individuals after the opening of a blood transfusion market. I show that these two representations of opportunities are linked to two different types of distinct preferences (altruistic or selfish). Depending on whether we value one type of preference or the other, the way we represent opportunities will differ, meaning that it is generally not possible to identify opportunities independently of the preferences we attribute to individuals.

Chapter 2. The argument advocating the need for the state to provide public goods consists of three distinct premises: (1) individuals face a public good problem; (2) there is no other way to solve it except by resorting to some form of coercion; (3) the improvement in everyone's situation resulting from coercive intervention is sufficient to compensate for the loss of control it implies. Economists generally have little to say about the third premise, even though it is a crucial part of the argument. Libertarians, who oppose state coercion as much as possible, may reject either premise (2) or premise (3) in the face of a public goods problem. Some of them have argued that premise (2) is false because so called insurance contracts can be voluntarily designed and implemented to coordinate individuals' contributions to the public good and produce it. I propose to consider what I call 'emergency situations', where these insurance contracts are not feasible due to the severity of these situations (for instance, the onset of a deadly epidemic or rapid climate warming). On what grounds could libertarians then reject premise (3) in such situations? I argue, following Sen, that libertarians are wrong to equate freedom and control. If freedom can exist where there is no control, coercive intervention may not be seen as depriving individuals of their freedom. If coercive intervention imposes nothing more on individuals than what they would have imposed on themselves if they had accepted an insurance contract to provide the public good (and assuming they would have accepted it if they could), their 'indirect liberty', as one might call it following Sen, is preserved. Libertarians who recognize the appeal of this notion of indirect liberty can accept premise (3) and the freedom-preserving character of some coercive state intervention.

Chapter 3. To explain phenomena, economists use a representation of the economic agent's behavior and objectives. Let us assume that individuals are indeed as the traditional representation supposes, which would mean

that positive economics actually captures the essential characteristics of economic behavior. Then, as highlighted by Hausman (2021) in particular, this would justify the reluctance of welfare economists towards paternalist interventions because paternalism would simply be suboptimal. Indeed, if (1) individuals choose what they prefer, and (2) they prefer what is best for them, then no paternalistic intervention, which would interfere with their choices, could make them better off. But if, on the contrary, individuals do not always, or not often, choose what is best for them (as suggested by some behavioral economists), the anti-paternalistic argument is refuted. I show how this objection, which has been put forward in light of the findings of behavioral economics, could have been formulated much earlier in the history of normative economics without giving up premises (1) and (2), because it is well known, ever since Thomas Schelling spoke about it, that in the context of strategic interactions, individuals do not, in a sense, choose what is best for them. This has significant implications for possible interventions and the definition of the scope of paternalism that have not been recognized: the possibility of what I call 'strategic paternalism'

Chapter 4. Anti-paternalism in economics involves respecting individuals' preferences, whatever they may be. This implies that there must be a coincidence between what matters to individuals and the evaluations that economists give of their situations. However, as I demonstrate in Chapter 4, this is not the case if we adopt a criterion of freedom or opportunity (as suggested by Sugden 2018a) to assess economic situations and at the same time admit the possibility that individuals have a preference for commitment. The chapter highlights this 'trilemma', which is a general incompatibility between anti-paternalism, the use of a freedom criterion, and the possibility of a preference for commitment. We can go further and show that an anti-paternalistic position implies attributing a minimal preference for freedom to individuals. Therefore, it is impossible to adopt a 'normative' anti-paternalistic position without having to make certain 'positive' assumptions about their behavior. To prove this, I consider the simplest possible model of interactions between a citizen and an economist providing recommendations to public authorities. The model shows that an anti-paternalistic position is consistent only if individuals are willing to be treated as such, which means that they value freedom, in a minimal sense.

Chapter 5. Commitment devices enable individuals to restrict their fu-

ture choices. What we could call 'hard commitment devices (HCDs) do so by making certain options materially costly, while 'soft' commitment devices (SCDs) achieve this by making them psychologically costly. The existence of markets for HCDs ensures that individuals with self-control problems can purchase them to prevent themselves from doing something tomorrow that they consider bad today. The arguments for or against markets for HCDs depend on a certain representation of the interests of individuals prone to self-control issues. Behavioral economists have proposed models of individuals composed of multiple selves with conflicting preferences, which have been used to advocate for paternalistic interventions. Sugden (2018a) criticized this view as being based on the unfounded idea that individuals should normally have consistent preferences and that their choices would be consistent if they did not have self-control issues, a claim that psychology cannot empirically support. According to Sugden, if we abandon this idea and instead admit that individuals are responsible agents who consider their past and future choices as their own, there would be nothing anomalous about thwarting one's own plans—such inconsistencies would simply be indications that people have changed their minds. However, the mere fact that individuals choose to use HCDs suggests that they see themselves as having multiple selves since they anticipate having preferences in conflict with those they have now. These two representations of the agent thus appear to be empirically unfounded and insufficient to justify normative conclusions. There is room for a different perspective that does not make such assumptions about the psychology of the agent. From the standpoint of Mill's liberty principle, according to which individuals should be left free to do what does not harm others, what matters is only whether this principle is respected or not, regardless of our representation of the interests of the agents. However, it can be shown that this principle cannot justify forcing individuals to keep their commitments to themselves. This conclusion would thus justify regulating a market for HCDs to limit the extent of binding commitments.

Chapter 1

Altruism and the Simple Argument for Markets

Many private goods fall into the category of 'contested commodities': (Radin 2001) blood, drugs, surrogacy, sex services, etc. A general argument for extending the sphere of the market to these goods appeals to the value of freedom: new markets give people more freedom to choose, whatever may be their preferences. By re-examining the debate between Titmuss and Arrow about the market for blood transfusions, I show that this simple argument is not conclusive because it fails to consider the motivations of 'impure altruists' (Andreoni 1990) whose preferences are denied by the creation of a market for blood. This objection offers a reinterpretation in economic terms of a disputed claim made by philosophers that the meaning of some goods (such as blood) changes when it becomes possible to exchange them on a market.

Introduction

Adam Smith, according to Friedman and Friedman (1990), 'analysed the way in which a market system could combine the freedom of individuals to pursue their own objectives with the extensive cooperation and collaboration needed in the economic field to produce our food, our clothing, our housing'. He showed that markets make individuals free to choose while at the same time delivering the goods. Both aspects are valuable, according to Friedman (1962/2002), since 'freedom in economic arrangement is itself a component of freedom broadly understood, (...) economic freedom is an end

in itself'. However, contemporary economics has focused almost exclusively on the second aspect, putting forward the efficiency of a market system and its merit for improving welfare. The fact that the Friedmans chose the title 'Free to Choose' for their popular book, relegating welfare in the background, suggests that this way of promoting markets is more in tune with common representations, or perhaps more likely to win support. An appeal to freedom of choice provides a simple and neat argument for expanding the sphere of the market. As Hausman et al. (2016, 94) sums it up in their introductory book on economics and ethics, 'markets permit the simple freedom of being able to choose among alternative as one pleases (provided that one has the means, of course)'.

The reference to this 'simple' freedom is the basis of what I will call the 'simple argument for markets'. This argument can be found virtually everywhere—everywhere else than in standard economic analysis[1]. Besides its role in philosophical and intellectual debates, it is also sometimes used in politics to convince people of the benefits of privatization[2], or at least to argue against state monopolies[3]. Its *a priori* nature makes it powerful and appealing to those, like the Friedmans or Sen, who value freedom for itself. But it is seldom the subject of analytic scrutiny[4].

I will consider in this chapter the value of this argument in justifying the creation of new markets. In substance, as Arrow (1972) formulated it, the possibility of performing new market transactions only adds to individuals' already existing alternatives—understood as mutually exclusive possibilities of actions—, leaving everything else unchanged. Indeed, people who are not interested in these transactions may abstain. Individuals' freedom is thus enhanced, as everyone has a bigger set of alternatives than before[5]. Because of his simplicity, the argument does not appear to rely on any empirical premise or substantive value commitment concerning the definition and measurement of the set of alternatives, or 'opportunity sets', as they are called in the lit-

[1]This has not always been the case, as Sen in particular has documented. 'The shift in the focus of attention of pro-market economics from freedom to utility have been achieved at some cost: the neglect of the central value of freedom itself.' (Sen 1999b, 27-28). See also Sen (1993).

[2]In Sweden, conservatives have extolled the virtues of a 'choice revolution' to convince citizens of the benefits of private provision of health care and social services in Sweden. See also references to the 'choice agenda', particularly in British politics.

[3]See Le Grand (2007; 2011).

[4]With the exception of Sen, as we will see.

[5]In the sense of the inclusion criterion, see section 3 for a definition.

erature devoted to the measurement of freedom. Besides, contrary to what is sometimes claimed (Herzog 2021), this 'simple' argument does not presuppose any belief in pre-established libertarian rights, as Nozick (1974) has described them—although it obviously presupposes the existence of private property, which is necessary to make markets work.

I will show that the argument falls short of this promise by discussing a particular (and often ignored) aspect of the famous debate between Arrow and Titmuss about the merits of a market for blood transfusion, compared to a voluntary donor system. Titmuss (1970/2018) pointed out that the institution of the market denies people 'the freedom to enter into gift relationships', to the disbelief of Arrow and many after him. A few philosophers (first Singer 1973, and then Anderson 1990, Radin 2001, Archard 2002) have tried to make sense of this perplexing argument by arguing that the opening of a market, altering not the mere possibility but the *meaning* of a blood donation, may deprive individuals of the opportunity to give 'the gift of life' to the recipient of the transfusion.

I will take a different route and stick to the standard perspective and modelling practice of economists, to show that the satisfaction of some particular kind of preferences—those of 'impure altruists' (Andreoni 1990)—is denied by the market for blood. This different route leads to the same destination, however: the opportunity to give 'the gift of life' which was previously accessible to individuals disappeared. As this contradicts the premise according to which the opening of a new market leaves everything unchanged in terms of freedom, the simple argument no longer holds. This discussion illustrates and extends Sugden's claim that the definition of opportunity sets always appeals to some particular set of preferences or value commitments—which means that it cannot be completely value-free (Sugden 2003). I will show that the simple argument only works if the preferences of impure altruists are disregarded, which means that contrary to what is often denied, markets may well impose a 'preordained pattern of value to which individuals must conform' (as Satz 2010 puts it) [6], even when these individuals refuse to transact.

In the first section, I provide a detailed analysis of the simple argument for markets, and its connection to the idea that markets are a value-free

[6] 'In a market system there is no preordained pattern of value to which individuals must conform' (Satz 2010, 23). The following will show in what way this is not true. See also Anderson's quotation in the first section.

space where anyone can express and develop their particular individuality, whatever may be their preferences. The second section presents Sugden's criticism of the claim that opportunity sets can be defined and measured without favouring implicitly some preferences or values. A simple principle for defining opportunities in real-world situations, which accords with the modelling practice of economists, is defined: according to the principle of relevance of value-differences, if some individual prefers some action to another one, then those two actions should be modelled as different opportunities accessible to him. The third section introduces the debate between Arrow and Titmuss about the 'right to give' and its reformulation by subsequent philosophers. The fourth section uses the principle defined earlier to show that the opening of a market for blood deprives individuals of a significant opportunity since the satisfaction of the preference of impure altruists is denied by the opening of a market for blood. I conclude by suggesting that markets are only amenable to what I call 'market-based' preferences, which are defined over the output of the market transactions—in terms of individual results and satisfaction—, neglecting preferences which also value contextual elements (as impure altruists' preferences do).

1.1 The simple argument for markets

The simple argument is easy to spot in speeches or discourses advocating free markets: simply put, markets make individuals free to choose. However, this formulation hides that there are two distinct ways in which an argument for markets can be made by relying on the value of freedom. Amartya Sen[7] deserves credit for making a crucial distinction between what I call the simple argument for markets, which does not involve any detailed analysis of how the market mechanism actually works, and does not pay attention to its outcomes, and the more elaborate argument that would need it. As Sen puts it:

> There is, of course, an obvious sense in which the freedom of transaction, an essential concomitant of the market mechanism, does directly make the parties involved more 'free to choose'. To

[7]This distinction is explored in Sen (1985; 1993; 1994; 1999). For an analysis, see Prendergast (2005; 2011).

> be able to transact freely, given other things, clearly does give the potential transactors more freedom to act (Sen 1994, 125).

According to the first, 'simple', argument, the very existence of the market mechanism involves the possibility for individuals to perform certain transactions associated with this market. Anyone who can afford to do their part of a market transaction—either by buying or selling goods and services—can in principle perform it, independently of their identity or motivations. The opening of a new market, which gives the possibility of performing new transactions in addition to those already available, thus enlarges the set of opportunities to act which are available to any individual and from which individuals may choose. Sen remarks that

> The claim is defendable (...) *without* the necessity of going into substantive analyses of how the market work, how the individuals substantively fare as a result of the transactions, how well off they end up being, or how much freedom to achieve they actually acquire through these means (Sen 1994, 126).

The distinction between 'freedom to act' and 'freedom to achieve', is reformulated in a later paper (Sen 1993) as a distinction between what Sen calls the 'process aspect of freedom', and the 'opportunity aspect of freedom'[8]. The former formulation, however, makes very clear that the opportunities to perform certain market transactions are only a means to something else, with which our 'freedom to achieve' is concerned—since performing a market transaction is not an achievement that someone would value for its own sake. By contrast, increasing one's wealth, being able to feed one's kids, etc. are achievements that one can expect to get by choosing the appropriate market transactions among those available to them. The 'simple argument', which only tells us that more market transactions means more opportunities to act, is silent on this aspect. It is much more difficult to state the conditions under which those expectations can be fulfilled because it depends on the outcomes of the market mechanism, which involves the participation of thousands of individuals in addition to the ones we are concerned with.

[8]The opportunity aspect of freedom is defined there as 'the capability to achieve', related to 'the real opportunities we have of achieving things that we can and do value (no matter what the process through which that achievement comes about)' whereas the process aspect of freedom lies in 'having the levers of control in one's own hands (no matter whether this enhances the actual opportunity of achieving our objectives)'

If market transactions are conceived as a means to achieve something else, there are many ways in which adding new possibilities of market transactions may prevent individuals from getting these achievements, and therefore lead to a decrease in their 'freedom to achieve': (1) so-called technological externalities, (2) pecuniary externalities and (3) other strategic effects. The first two cases are well known: for example, if the production of some new good involves some sort of pollution, it may be that some people end up poorer as a result (pollution decreases the value of the land they own, for example). The opening of a new market may also affect the price of the only good that someone was endowed with, thus making him poorer. The third case is not necessarily connected to the functioning of a market. It has to do with the fact that having means or resources ready to use can be detrimental to someone in a strategic context. Suppose that a new market for kidneys has recently opened. This enables thugs to threaten a poor family man and make him sell his kidney to comply with the threat. Before the opening of the market, the family man had no opportunity to gain as much money and was therefore not a target for the thugs. Adding the possibility of selling his kidney has thus worsened the prospects of this individual, in terms of his freedom to achieve a decent life with (or without) both kidneys[9].

All these mechanisms by which someone's 'freedom to achieve' may be thwarted involve some kind of side effects, but—especially in the case of (2)—these are not avoidable. Unless we can be certain that they would not come into play, it is not possible, as noted by Sen 'to jump to an immediate conclusion that, for the people as a whole, market transactions must expand substantive freedoms that we actually enjoy'. But this is not the end of the story. Market transactions exist as means to some possible ends. The fact that, under certain conditions, using these means may be detrimental for some people because of side effects does not contradict the fact that it is, in a sense, good to have these means. It is precisely because he is now solvable—by being able to sell his kidney—that the family man described earlier is vulnerable to threats from the thugs. And negative external effects (either pecuniary or technological) happen *in spite* of the gains from exchange obtained by the individual performing the transaction. Even the fact that some people are not able to afford to perform the transaction, because they do not possess the good which is sold or cannot pay the price to buy it, does

[9]The case for closing a market for kidneys could thus be an instance of strategic paternalism, see chapter 3 on these issues.

not disqualify it as being a valuable opportunity, although perhaps not a 'real' or 'substantive' one, as Sen puts it[10]

Indeed, the point of having the possibility to do market transactions is that they are general means to reach certain ends that individuals might have, whatever may be these ends. Individuals can access goods and services, or get paid for what they offer, independently of who they are and what they value. And if they do not value what markets have to offer, they are free to pass up the opportunity to buy or to sell. As Elizabeth Anderson, who is not an advocate of market 'imperialism', recognizes:

> The market provides individual freedom from the value judgments of others. It does not reward any one individual's preferences as less worthy of satisfaction than anyone else's, as long as one can pay for one's own satisfaction. (Anderson 1990, 183)

The idea that markets are value-free zones where everyone can get what they want, whatever may be their preferences, makes them particularly amenable to serve as 'a space for the development and expression of individuality, as Sugden puts it. In the liberal tradition associated with John Stuart Mill, having more opportunities is good in itself insofar as it promotes the free expression of individuality. More precisely, having more opportunities is good for three related reasons: it helps people (1) to shape their lives through their choices—according to their own judgements and desires—, (2) to develop their faculties of judgement in the process of choosing; (3) letting people do (1) and (2) presents the world with a variety of 'experiments of living' from which other individuals can learn[11]. But if the opportunities offered to individuals come attached to some particular set of values, or must have a 'reasonable' character, they would not serve as a 'space for individuality', as Mill and Sugden envision it. The full expression of individuality would not be achieved, since the set of opportunities available would

[10] As emphasized by Binder (2021b) since the arguments about the ability of markets to promote freedom crucially depend on the conception of freedom which one invokes, some clarifications are needed. MacCallum's 'formula of freedom' states that agent X is free from a set of relevant constraints Y to do, or not do, Z. Here, Xs are individuals, and the set of constraints involve legal or moral constraints (such as the 'repugnance' of certain transactions, as Roth 2007 puts it) but not economic constraints such as a lack of money, and Z is the performance of some actions. Sen (1993) frame the simple arguments in terms of negative freedom (which he calls 'immunity from encroachment') but does not specify clearly what is the Z variable.

[11] see Sugden (2003) for elaboration on this.

be restricted to what authorities or majorities deemed as 'reasonable', or 'valuable'—excluding some possible ways of valuing the goods and services which individuals may have[12]. Since market opportunities would not be of this sort, they would be valuable as constituents of the most extensive 'space for individuality' that the liberal tradition can uphold. Thus, a crucial part of the simple argument for markets is that they provide individuals with more opportunities to act that are valuable to them, whatever may be their preferences.

How can we know that every individual really has more of these opportunities? An argument that concludes that some intervention or social change provides more freedom to all individuals has to compare the opportunity sets of every individual before and after the intervention. How should this comparison be done? The freedom of choice literature proposes a number of criteria for measuring the freedom that a given opportunity set can provide. A simple criterion, which seems appealing despite its limitations[13], is the inclusion criterion. According to this criterion, we can conclude that all individuals have more freedom than before if new opportunities are open to them after the intervention, while all the opportunities available before are still open. We thus have only two conditions to check. The opening of a new market really gives new opportunities for all individuals, understood as possibilities to perform some particular market transactions. It is essentially a regulatory change that does not seem to prevent people from doing anything—individuals not interested in newly available market transactions may abstain. It would thus seem that all previous opportunities to act are kept open by the opening of a new market. I will show in the following that this is not necessarily true, which prevents the argument from being conclusive. The simple argument may therefore be expressed under this form:

1. Individuals have more freedom to act if everyone is provided with new opportunities to act (valuable whatever may be their preferences) while

[12]This echoes Friedman's argument contrasting political 'conformity' and market 'diversity' (Friedman 1962/2002).

[13]This criterion has been criticized by Sen (1991), in particular, because it implies that any new opportunity would enhance someone's freedom, which would not be true for Sen if this opportunity is seen as bad by the agent (such as the possibility to be beheaded at dawn, to use Sen's example). But if we are concerned with creating a 'space for individuality', we should not base our judgement on whether new possibilities should count as genuine opportunities on a particular set of individual preferences. See Sugden (2018a) for a defence of an inclusion criterion along these lines.

the ones they already have are kept open.

2. The opening of a new market always gives every individual new possibilities to perform some market transactions, which are new opportunities to act (valuable whatever may be their preferences), while keeping every opportunity to act that they had before open.

3. The opening of a new market gives more freedom to act to individuals.

What will be challenged in the following is premise (2), on the basis that the opening of a market sometimes fails to keep some opportunities to act previously available open, if the opportunities to act deemed relevant for the comparison are understood to be valuable whatever may be our preferences. This argument—if it was valid—would give what Sen calls a '*prima facie* presumption in favour of allowing people to transact as they like' (Sen 1999b, 26). It would only be a presumption since it might be that the market mechanism results in limiting the overall 'freedom to achieve' of individuals—and in that case, it might be preferable to avoid opening the market. But it would be a powerful presumption since it is applicable everywhere, for any kind of market, and one that is not easy to rebut if we value the 'space for individuality' I described in this section. As Fabienne Peter summarizes: 'Markets appear not only as mechanisms that efficiently allocate resources but (...) as systems that automatically legitimize themselves' (Peter 2004, 3).

1.2 The problem with opportunity 'as a space for individuality'

To check if premise (2) is true in certain contexts, we would need to build what Beck and Jahn (2021) call a 'normative model', which would enable us to judge whether or not all previous opportunities to act are preserved after the opening of a new market and thus whether the argument is conclusive. For Beck and Jahn, a normative model 'aims at providing normative guidance to agents' and

> the way in which normative models exert normative guidance is by means of extending normative justification to cases of normative uncertainty (i.e., a situation in which we are unsure about which action we should perform). (...) This function consists

> of two elements (...). First, normative models allow us to summarize normative verdicts that are justified independently of the model. They do so by means of a (typically) sparse set of idealizing assumptions that entail these verdicts. (...) The second element (...) consists in the fact that normative models allow us to project the identified pattern onto novel situations (Beck and Jahn 2021, 142).

Whereas normative arguments (such as the one presented earlier) rely on their premises, normative models rely (just as descriptive models) on assumptions about agents and their environments. These assumptions help us to extend the normative 'verdict' or conclusion already obtained to situations where the normative judgement is not yet settled. I will take for granted that the conclusion of the simple argument is valid when it comes to ordinary markets, where basic consumption goods are exchanged on the basis of their ability to satisfy each seller's own preferences. But as we will see in the next section, the debate between Arrow and Titmuss about the market for blood shows that we here have a case of 'normative uncertainty': we are unsure if an intervention allowing the selling of blood would really give more freedom to everyone. A normative model can help to determine whether the same patterns that we observe in the market for basic consumption goods can also be found in the market for blood.

The normative models that I will describe in this section are the ones involved in the 'pure quantity' approach to measuring freedom, as Sugden (2003) describes it. This approach defines various ways of measuring the freedom of individuals (for example by a discrete 'cardinality' rule, by a spatial rule, or by diversity criteria) in a purely quantitative manner, without any consideration for their preferences. As we have seen in the previous section, the simple argument for markets relies on the fact that market transactions are opportunities to act which are valuable whatever may be the preferences of individuals. As emphasized by Sugden, the 'pure quantity' approach would be the only one suitable to build a normative model reflecting the value-free perspective of measuring opportunity 'as a space for individuality' and thus also to provide a basis for the simple argument. In particular, we would expect that a pure quantity approach counts as genuine and valuable an opportunity even if it is deemed 'inferior' or 'unreasonable' according to other approaches[14], since it only takes into account the sheer existence of opportu-

[14]See Pattanaik (1998) for a definition of a ranking in terms of freedom mentioning

nities and not their value to the individual. However, according to Sugden, this pure quantity approach fails to be completely value-free. A value judgement is required to decide what aspects of the world counts (or not) as a genuine opportunity—even if we do not take into account individual values in the measurement of freedom.

Indeed, in normative models designed to measure freedom, values may enter the analysis at two distinct levels. The modeller first needs to identify the set of all opportunities accessible to the individual whose freedom is measured, and secondly, to decide to use one ranking rule or another to measure the freedom this 'opportunity set' would give to the individual. In this second stage, the modeller can make his value judgements explicit: the freedom of choice literature is very transparent on what information about individual preferences or 'reasonable preferences' are used to measure freedom. But this is not so easy in the first stage when deciding what aspects of the economic world can count as opportunities open to individuals. In the freedom of choice literature, opportunities are often considered as given. There are very few guidelines on how to map aspects of real-world situations to opportunity sets. In any case, these modelling decisions must be made through the use of idealizing assumptions which are not necessarily obvious to the modeller herself. As Sugden points out,

> The problem is this: in order to measure opportunity in a real-world situation, we have to be able to say whether two putative options should be treated as distinct or to be able to specify how significant the difference is between one option and another. That requires us to locate options in some conceptual space in which relations of similarity and difference can be defined. But there are many such spaces, none of which is uniquely privileged. If we try to resolve this problem by appealing to an intuitive understanding of opportunity, we are drawn towards concepts of similarity and difference that refer to reasonable or normal preferences. I conclude that the search for a non-arbitrary, pure-quantity measure of opportunity cannot succeed. (Sugden 2003, 803)

What would be wrong with the simple argument, if we follow this criticism of the pure quantity approach to measuring freedom? The importance of the

reasonable preferences.

pure quantity approach lies in its ability to be an all-encompassing measure of freedom, allowing for all the ways in which an individual can live his life. The trivial fact that aspects of the world can be discretized in a considerable number of ways (and that this discretization is somehow arbitrary) does seem to be a major problem for the pure quantity approach—as it is committed to avoiding *all* value judgements because they would lead to disregard some non-conformist way of living one's life. To make Sugden's criticism more precise, I will define a simple requirement that any pure quantity modeller should respect. Let us grant that possible actions that an individual can perform may be discretized in a considerable number of ways and that we have to choose one to produce a normative model of opportunities. To identify the opportunities to act that we want to put into our pure quantity model for measuring freedom, we should, in any case, apply the following principle:

Relevance of Value Differences: for any couple of possible actions that an individual can perform, if this individual values these two possible actions differently, then the modeller should represent these two actions as distinct opportunities to act in her opportunity set.

This principle states a necessary condition for identifying opportunities to act in a real-world situation from a pure quantity perspective. If the individual himself treats two actions as distinct, as he values more the performance of one action than the other, the modeller should also treat these actions as distinct. This modelling principle reflects the practice of economists (either positive or normative), who do not discriminate between preferences and attempt to track every trade-off that individuals may face in their lives (and therefore the differences in valuations that caused it)[15]. Since the goal of the pure quantity modeller is to provide a representation of the opportunities opened to the individual that enables him to express whatever preferences he may have, the least that the modeller can do is to defer to the individual's own representation of his opportunities. I will not discuss how in practice such a discretization of actions should be done since my goal is simply to describe a case that falsifies premise (2) of the simple argument. As we will see, it is enough to use this principle to show that the simple argument is

[15]This may be truer of the new brand of behavioural economics, which makes room for a wide range of possible motivations, than of old welfare economics which would propose to limit itself to 'man's conduct in the business part of this life', as Marshall (1890/2013) puts it.

invalid.

Since the principle of relevance of value differences is a necessary condition for identifying opportunities from a pure quantity perspective, it has to be respected each time a representation of the opportunities opened to an individual is given. If that is not the case—because two possible actions that have distinct values for some individuals are treated as one—, this representation of individuals' opportunities fails to reflect a possible representation of their own opportunities. We could thus conclude that the representation of individuals' opportunities by the modeller cannot be used to measure the extent of the 'space for individuality' opened to them, because of its absence of neutrality with regard to all possible valuations of the situation: some particular values or preferences are set aside in this process of representation.

1.3 Arrow, Titmuss and the market for blood

In his book *The Gift Relationship* (1970/2018), Richard Titmuss compared systems of blood donation in the United Kingdom, which only relied on voluntary donors, and in the United States, where blood supply was provided by for-profit companies which remunerate donors. As Fontaine explains, 'blood, so crucial to body integrity, was ideally suited for illustrating the centrality of gift-giving to the maintenance of the body politic' (Fontaine 2002, 404), which was the concern of Titmuss. Titmuss claimed that letting market mechanisms operate to supply blood had disastrous effects. He claimed, in particular, that it is inefficient, as the quantity and quality of the blood donated is inferior under a market system compared to a system where only voluntary donors are allowed to give. Concerning quality, compared to voluntary donors, paid donors are less truthful and less likely to reveal a full medical history and provide information about contacts with infectious diseases. Concerning quantity, voluntary donors tend to be less motivated to give if their donation is accompanied by a monetary reward, even if they can choose not to receive it. This latter inefficiency has to do with what has been called the 'crowding out' of the intrinsic motivation to give[16], which is not my concern here since it does not touch the simple argument for markets. Interestingly, Titmuss also said that the opening of a market for blood 'represses the expression of altruism, erodes the sense of community' and that

[16]The term seems to have been coined by Frey and Oberholzer-Gee (1997), see Gneezy, Meier and Rey-Biel (2011) and Bowles (2016) for a discussion.

it 'limits both personal and professional freedoms', which goes beyond pointing at inefficiencies (Titmuss 1970/2018, 210). This latter claim contradict directly the simple argument, as we will see. Titmuss's study presents itself as a demonstration of the failure of economic analysis and Arrow (1972) answered it. This answer was in turn discussed by Peter Singer (1973) in the same journal. Since then, several papers have discussed the argument made by Titmuss that a significant and valuable 'right to give' is erased by the mere existence of market transactions, most often to dismiss it (Lomasky 1983, Dworkin 1982).

On the contrary, I will show in the next section how this argument can make sense, even from the point of view of economic analysis, and in particular how a pure quantity normative modeller would have to take it into account. According to Titmuss, 'policy and processes should enable men to be free to choose to give to unnamed strangers. They should not be coerced or constrained by the market' (Titmuss 1970/2018, 206). Later, Titmuss argues that 'if it is accepted that a man has a social and biological need to help, then to deny him opportunities to express this need is to deny him the freedom to enter into gift relationships' (ibid., 207). This is puzzling, since, as proponents of the simple argument would argue, the opening of a new market for blood does not prevent individuals from giving their blood voluntarily, just as they did before. As Arrow puts it,

> Economists typically take for granted that since the creation of a market increases the individual's area of choice, it therefore leads to higher benefits. Thus, if to a voluntary blood donor system we add the possibility of selling blood, we have only expanded the individual's range of alternatives. If he derives satisfaction from giving, it is argued, he can still give, and nothing has been done to impair that right. (Arrow 1972, 349-350)

This corresponds exactly to the formulation of the simple argument, adapted to the case of the 'gift relationship'. The opening of a market for blood gives people the possibility to perform some market transactions where they give their blood in exchange for payment while preserving the opportunity to give their blood for free. How can anyone's 'freedom to enter into gift relationships' be denied by the opening of a market for blood? It is as if, for Titmuss, the opportunity to give blood voluntarily, in an altruistic manner, did not exist anymore. Singer rephrases Titmuss's argument as follows:

> Titmuss's idea that the creation of a commercial system threatens the right to give is not so much mistaken as inadequately developed. The right that Titmuss saw threatened is not the simple right to give, but the right to give (...) something that cannot be bought, that has no cash value and must be given freely if it is to be obtained at all. This right, if it is a right (it would be better to say 'this freedom') really is incompatible with the freedom to sell, and we cannot avoid denying one of these freedoms when we grant the other[17]. (Singer 1977, 163-164)

It is a trivial truth that giving people the opportunity to sell a good on a market also takes away their opportunity to give it as 'something that cannot be bought'. This reformulation makes Titmuss's argument formally correct since the opening of a market for some good fails to keep open the opportunity to give 'something that cannot be bought'. We thus have two possible and incompatible representations of the opportunities available to individuals:

- According to Arrow's representation, after the opening of the market, individuals have a (new) opportunity to sell their blood in addition to the opportunity to give it freely, which was available before and still is after.

- According to Titmuss and Singer's representation, after the opening of the market, individuals have a (new) opportunity to sell their blood, but they do not have the opportunity to give it as 'something that cannot be bought' anymore. On the other hand, they still have the opportunity to give it without receiving a payment.

This second representation formally contradicts premise (2) of the simple argument, since some opportunity previously available (the opportunity to give something that cannot be bought) has disappeared as a result of the opening of the market. If some opportunity to act is lost, for some individuals, because a market has opened, the simple argument is refuted. The question thus becomes: is the opportunity to give blood really the *same* before and after the opening of the market? Has the opening of the market really *changed* the nature of the opportunity to give? It is the answer to

[17]Quoted by Lomasky (1983, 252-253).

these crucial questions that separate the two representations. If the second representation is the correct one because the opportunity to give what cannot be bought reflects a genuine way of expressing one's individuality—and one which cannot be preserved by the opening of the market—, the simple argument no longer holds.

But not everyone agrees that this opportunity to give what cannot be bought makes any sense at all. According to Lomasky, it

> *conceptually includes* the non-performance of certain actions by others. This might be called an *inherently exclusionary* liberty—or so it could be called if this marked out a genuine liberty at all (...) It is perverse to claim a liberty to be first in a race, to make more money than anyone else, to be the most lavishly rewarded scholar in one's discipline, to enjoy an enforced monopoly. Each is more properly described as the claim of a privilege[18]. (Lomasky 1983, 258)

Lomasky seems to hesitate between two different criticisms: either the opportunity to give what cannot be bought cannot be counted as a 'genuine' opportunity at all, or it can, but should not be taken seriously as it would be 'perverse' to claim a 'liberty' which is, in essence, a privilege. In any case, it is worth noting that, in a social context, most actions involve the non-performance of certain actions by others. The associated opportunities, such as parking one's car in the street (thus preventing other people from using the space), are not necessarily seen as privileges, and they may be considered genuine opportunities by the individuals who have and value them. Lomasky's argument seems too far-reaching. It would rule out the possibility of representing opportunities in a social context. From the neutral perspective of a pure quantity approach, the question should not be whether some opportunity is 'exclusionary' or not, but rather whether individuals *could* see it as meaningful. Since the principle of relevance of value differences is a necessary condition for identifying opportunities from a pure quantity perspective, it has to be respected each time a representation of the opportunities opened to an individual is given. If that is not the case—because two possible actions that have distinct values for some individuals are treated as one—, this representation of individuals' opportunities fails to reflect the

[18]Emphasis is not mine.

way they see their own opportunities. We could thus conclude that the representation of individuals' opportunities by the modeller cannot be used to measure the extent of the 'space for individuality' opened to them, because of its absence of neutrality with regard to all possible valuations of the situation: some particular values or preferences would be set aside in this process of representation.

1.4 Impure altruism and the gift of life

In light of the principle of relevance of value differences stated before, two actions that are valued differently by someone must be represented as two distinct genuine opportunities by the modeller. If people really value the opportunity to give what cannot be bought differently than the opportunity to give when it can be bought, they should be represented as distinct opportunities, whether or not the latter opportunity is seen as a privilege. The question then becomes whether this difference between giving what cannot be bought and giving when it can really matters to individuals. This is indeed the case, if we follow Arrow himself, as well as the more recent insights from positive economics about 'impure altruism'—in other words, Arrow give in his own text the reason why the simple argument, which he also employs, is not valid. Arrow presents two distinct kinds of altruistic motivations for giving blood. According to the first, what matters to individuals giving blood is the welfare of others. According to the second, what matters to individuals giving blood is (also) their own contribution to the well-being of others:

> (1) The welfare of each individual will depend both on his own satisfaction and on the satisfactions obtained by others. We here have in mind a positive relation, one of altruism rather than envy.
>
> (2) The welfare of each individual depends not only on the utilities of himself and others but also on his contributions to the utilities of others. (Arrow 1972, 348)

Following James Andreoni's (1990) work on altruism we may call the first kind of motivation 'pure altruism', and the second 'impure altruism'. It is also better known under the expression of 'warm-glow giving'. The mention of this second motivation is due to the intuitive observation that 'humans (...) enjoy gratitude and recognition, they enjoy making someone else happy

and they feel relieved from guilt when they become a giver'. This implies that 'people are not indifferent to their own voluntary gifts and the gifts of others, they strictly prefer, all else equal, that the gift come from themselves' (Andreoni 2006).

I will be more precise: what matters to the particular 'impure altruist' I consider is the difference that his action makes in the welfare of someone else. Giving a small amount of money to a homeless man would make someone feel good because they know it will significantly increase their welfare. Giving the same amount to a billionaire or even to one of their colleagues would be less rewarding since it would not significantly increase their welfare. They would thus value differently the two cases, even if the amount of money that they gave is the same. And, crucially, I will take that the value that their contribution has for them depends on the fact that it is *causal* in increasing the recipient's welfare. Suppose that if they had not given this amount to the homeless man, any one of the passers-by would have done it in an instant. In that case, the homeless man has no particular reason to be grateful to them, since, as goes the saying, 'anyone would have done it'. The mere fact that we need to point that out in a show of modesty (knowing that it is not true) suggests that the causal impact of our contribution is significant for us and that it is what gives us a 'warm-glow' feeling.

In the framework of Andreoni (1990), a charitable donation is modelled as a public good. Here, following Arrow's suggestion, I will take the utility of the recipient of the donation as the public good, which will enter the utility function of the donor himself. The utility function of a pure altruist donor depends only, as Arrow indicates, on his own consumption and on the utility level of the recipient of the donation. The utility function of an impure altruist depends, in addition to the above, on his contribution to the welfare of the recipient. This means that to measure the utility that the donor derives from giving, we need to determine the counterfactual level of utility of the recipient if the donor had not given and use the difference with the actual utility level of the recipient as an indicator of the extent of his contribution.

To make things as simple as possible, I will consider only the two extreme cases where the donor is either a pure altruist or an impure altruist who only cares about this contribution to the welfare of the recipient (and not about his actual welfare). I will also suppose that both utility functions are additive with respect to their arguments and that the level of (private) consumption of the donor is given. Let us also suppose that the individual whose welfare

I will consider is the only person who is willing to give to the donor (but not the only one willing to sell). This is obviously not very realistic, but sufficient for my purpose, which is only conceptual.

In all the situations which I will describe, the recipient may have three different utility levels. I will call v_1 the utility of the recipient when he receives the blood transfusion without having to pay for it, v_2 the utility of the recipient when he receives the blood transfusion and has to pay for it, and v_3 the utility of the recipient if he doesn't receive any blood transfusion. We have of course $v_1 > v_2 > v_3$. Even if it is not necessary for my argument, we could suppose that the utility of the recipient is cardinal and that the difference $v_1 - v_2$ is much lower than the difference $v_1 - v_3$, since the recipient cares a lot more about his life than about his money. From the point of view of the pure altruist donor, what matters is only the level of utility of the recipient once he has given, which is the same amount v_1 when the market for blood exists and when it does not. The increase in the utility level of the donor that he gets by giving is the same whether or not there is a market or not. In other words, the utility of the pure altruist is left completely unchanged by the opening of the market, because he only cares about the welfare of the recipient. Note also that the utility of the selfish person who gives nothing is completely unchanged in both situations. This reflects Arrow's statement, as quoted in the previous section: 'if he derives satisfaction from giving, he can still give, and nothing has been done to impair this right'.

Now consider the case of the 'extreme' impure altruist. When only giving is possible because the market for blood has not opened, his contribution is measured by the difference between the utility level of the recipient when he has received the transfusion, and his utility level without it, which is $v_1 - v_3$ because the recipient cannot pay to receive a transfusion and would be much worse off without it. This difference reflects the causal impact of a donation where there is no market for blood[19]. The increase in the utility level of the donor is therefore $u(v_1 - v_3)$ (where u is the component of his utility function affected by his contribution to the recipient's welfare) and it

[19]In real situations, there are obviously several donors, and the causal contribution of a particular donor, as I have defined it, would be nonexistent if the recipient could find another donor right away. It must be assumed that they value their causal contribution as a group, not as a particular donor. In any case, it seems plausible to say that actual donors would feel that they contribute more to the welfare of the recipient in a context where there is no market for blood.

cannot be larger. When the market opens, the impure altruist is worse off because his contribution is lessened by the fact that the recipient can always access the market and pay for a blood transfusion. His contribution is now the smaller difference $v_1 - v_2$. Since $u(v_1 - v_2)$ is smaller than $u(v_1 - v_3)$, Arrow's statement is falsified: the 'right' to give of the impure altruist has been impaired, as he derives satisfaction from the causal impact he has on the welfare of the recipient, and not from the welfare of the recipient in itself.

Applying the principle of relevance of value differences, we can conclude that if individuals are selfish or pure altruists, we have no reason to treat as a genuine opportunity the opportunity to give what cannot be bought, since these individuals value exactly in the same way the possibility of giving before and after the opening of the market. But this is not true of the impure altruist, as I have described it. The opportunity to give what cannot be bought has considerably more value to him than the opportunity to give when it can be bought because he cannot make in the latter case a contribution as large (in terms of causal impact) as in the former case. The opening of the market deprives him of something: the possibility to make a significant contribution to enhancing the life of someone. The conclusion is that premise (2) of the simple argument no longer holds in this case, since at least one significant opportunity is lost with the opening of a market for blood, valuable to all the impure altruists. If impure altruism is a genuine way to express one's individuality, the simple argument is not conclusive.

At this point, we might ask if it is really appropriate to take into account the motivation of the impure altruist when evaluating the simple argument[20]. After all, the impure altruist is not really an altruist if he only values his (causal) contribution to the welfare of others because it gives him a 'warm glow'. But his preferences are other-regarding even if he is not an altruist. Why should we include him? The answer is that there may be more in the opportunity to enhance the life of someone by one's own actions than the warm glow that one would get out of it. As Titmuss has emphasized, it may make sense to give people this opportunity not only because it would make them feel good, but also because it could foster feelings of reciprocity between fellow citizens and improve their trust in each other[21]. The vitality

[20]Brennan and Jaworski (2016) would argue that impure altruists are not 'entitled' to have the opportunity to 'give what cannot be bought' and Lomasky (1993), as we have seen, that it is 'perverse'.

[21]'In not asking for or expecting any payment of money, these donors signalled their belief in the willingness of other men to act altruistically in the future and to join together

of the debate that went on after the publication of Titmuss's book about the effect of markets on the expression of gift-giving behaviour shows[22] that there is no pressing reason to 'launder' impure altruists' preferences when evaluating individual situations (as Goodin 1986 puts it). The opportunity to make a difference in someone's else life is valuable as a signal of trust and a call for an impersonal kind of reciprocity. As Steiner (2015) points out, the gift relationship described and praised by Titmuss is very distinct from that described by Mauss (1925/2023) in his anthropological study about giving in 'archaic' societies. In the latter case, the gift relationship is subject to three obligations: to give, to receive, and to give back. By contrast, in the case of what Steiner calls the 'organizational gift'—of which blood donation is an example—the two latter obligations do not exist. For Steiner, 'in the organizational gift, there is no similarity between both individuals, connected only the by the willingness of one of them to help a suffering stranger'. The opportunity to help a stranger, independently of her personal relation to us, is precisely what the organization of blood donations provide, and what the impure altruist values.

The question raised here touches on a recent debate in philosophy, about 'commodification', and the 'moral limits of the market'. Commodification can be defined as the act of 'allowing certain things to be for sale' (Brennan and Jaworski 2016, 19). Should blood transfusions be 'commodified'? Opponents of 'market imperialism', such as Anderson (1990), Radin (2001), and Sandel (2012) have taken up the arguments of Titmuss and Singer and reformulated them into what may be called a 'semiotic argument' (in the terminology of Brennan and Jaworski). According to this argument, allowing for the selling and buying of blood alters and corrupts the *meaning* of blood donation. This would be the reason why the 'freedom to enter into gift relationships' emphasized by Titmuss is undermined by the commodification of blood transfusion. Consider how Anderson presents this change in meaning: 'the significance of my volunteer donation is trivialized when other blood is paid for. If blood is also a commodity, then all I have given to the recipient is the cash equivalent of the blood, not the gift of life itself' (Anderson 1990, 198). As explained by Radin, this would have the consequence that 'altruism is foreclosed if both donations and sale are permitted' (Radin, 2001, 96).

to make a gift freely available should they have a need for it. By expressing confidence in the behaviour of future unknown strangers, they were thus denying the Hobbesian thesis that men are devoid of any distinctively moral sense.' (Titmuss 1970:236)

[22]See Fontaine (2002) for an overview.

Behind this change in meaning—from 'the gift of life' to the 'cash equivalent of the blood'—lies the differences between receiving a free blood transfusion and death (or risk of death), on one hand, and between receiving a free transfusion and paying for it on the other hand. These differences are reflected in the utility levels $v_1 - v_3$ and $v_1 - v_2$, which capture as a change in valuation the change in meaning referred to by Anderson. There is therefore a close connection between Anderson's emphasis on the 'trivialization' of voluntary blood donation and the preference structure of the impure altruist. Who would suffer from the 'trivialization' of voluntary blood donation? If the answer is: no one, opposition to commodification risks being branded as paternalistic. But under the representation of opportunities that I just described, it appears that impure altruists are, by definition, sensitive to this trivialization, as it affects the prospect they have of making a difference in some stranger's life by their donation. The impure altruist would prefer to make the gift of life and save someone's life rather than giving the cash equivalent of blood. The meaning of this change in meaning, so to speak, could be that the commodification of blood transfusions prevents all of us from being what we could (and, according to Titmuss, should) be—that is, impure altruists who fulfill their aspiration.

Anderson's 'semiotic argument' against markets can thus be reformulated as an externality argument. By selling their blood, suppliers of the market for blood are having an *immediate* negative external effect on the welfare of impure altruists. Giving is no longer valuable for them, and their welfare decreases. This negative external effect must imply—according to the principle of relevance of value differences—that the possibility of giving what cannot be bought, however 'exclusionary' it is, is valuable and therefore a genuine opportunity. And this opportunity is lost as a result of the opening of a market for blood, whether or not there are impure altruists that are affected by this opening. This possibility may have been recognized by Sen, who wrote: 'In a competitive market, the levers of decision and control are in the hands of the respective individuals, and in the absence of particular types of externalities (dealing with the control of decisions), they are left free to operate them as they choose' (Sen 1993, 527). This qualification of the simple argument suggests that for Sen, externalities are not limited to impacts on the *welfare* of individuals. We could also speak of externalities related to the control of decisions by individuals, and, presumably, to their freedom to act. If this reading is correct, Sen could acknowledge that individuals are deprived of a genuine opportunity when a market for blood is opened, as

they no longer can decide whether or not they will give the gift of life.

Conclusion: the problem with the simple argument

According to the simple argument, the opening of a new market always gives people more freedom to act, but not necessarily more freedom to achieve. As new market transactions are available, more actions are available to individuals, while (apparently) no opportunity to act is removed, even for those who are not interested in the transaction or find it repugnant. What is crucial to this argument is the fact that the opportunities considered are valuable from any perspective, with regard to any kind of preferences. The market is supposed to take individuals' preferences as they are. And yet, giving individuals the opportunity to sell blood generates an external effect on impure altruists, as they no longer can make a difference in someone else's life as large as before. The opportunity to sell blood implies that the opportunity to give the gift of life, as Anderson would say, no longer exists. This is precisely what impure altruists value the most, and what they have lost. It is thus incorrect to say that the opening of a new market leaves every preexisting opportunity to act as it is: valuable opportunities to act have disappeared as a consequence.

It is not only, in Sen's terms, the 'freedom to achieve' their life-saving goal that is taken away from impure altruists, but the mere 'freedom to act' according to their preferences, since the opportunity to give the gift of life—which is instrumental in saving others' lives—has disappeared. Note also that this argument does not assume that any particular cultural or religious meaning is given to the goods. It is not because blood would have a cultural or religious significance as something that should not be bought that impure altruists have lost an opportunity. It is only because of the simple fact that blood can save someone's life, and make a huge difference in his welfare, which impure altruists value.

What remains, then, of the simple argument? There are certain preferences that the market cannot serve. But what is relevant here is that it destroys the possibility that some preferences—which are respectable as a way to express one's individuality—could be ever satisfied as a result of the opening of a market. The lesson is that the simple argument is somewhat cir-

cular: if it gives more opportunity to act for everyone, it is only with regard to these opportunities that are associated with what we may call 'market-based' preferences. Let us say that preferences are market-based when their satisfaction does not depend on elements of the broader social context, as the preferences of the impure altruist do. Once again, there is something to learn from Titmuss about this: 'choice cannot be abstracted from its social context, its values and disvalues, and measured in "value-free" forms. Blood distribution systems cannot be treated as autonomous independent processes' (Titmuss 1970/2018, 208).

The preferences of a selfish individual are market-based in this sense, and this is also true of the preferences of a pure altruist. They only care about the final level of satisfaction attained by individuals, independently of the broader process by which individuals came to be satisfied[23]. The range of choices opened to them can therefore be 'abstracted from its social context', but this is not the case for preferences which are structured differently, such as those of the impure altruist. Market freedom, in other words, can only accommodate certain types of preferences, at the expense of others. The rhetoric of the free choice provided by markets cannot be backed by an argument as sweeping as the simple argument for markets would seem to be. It can only become sound if the context in which market transactions take place—and the possible third-party effects influencing people's achievements—are given serious consideration.

[23]This relates to the 'process aspect of freedom' as defined by Sen (see Sen 1993). In Sen's terms, it turns out that the opening of a market does not always improve freedom in terms of its 'process aspect'.

Chapter 2

Preserving Freedom in Times of Urgency

The chapter shows how the provision of public goods through a public intervention forcing individuals to contribute can be said, paradoxically, to preserve their freedom—when they face what I call a 'situation of urgency'. By this I mean a situation in which cooperation between individuals is needed to address a catastrophic situation but is costly for individuals, such as the start of a serious epidemic, an imminent invasion by a foreign power or rapidly evolving climate warming. It reconciles a libertarian framework centred on rights, inspired by Nozick, with public coercive interventions meant to avoid severe collective losses. It concludes that, contrary to what is often claimed, measures such as the imposition of a lockdown, conscription or strict quotas on carbon-intensive consumption, are not necessarily liberticide.

Introduction

The inability of markets to provide public goods at a level sufficient to ensure efficiency is one of the best-known and most frequently highlighted market failures. The financing of public goods by voluntary contributions, often modelled as a prisoner's dilemma, fails to produce a Pareto optimum. This would justify the use of coercion by public authorities, forcing individuals to contribute to finance or produce an efficient level of public good. The call for coercion is justified, within traditional welfare economics, in purely *welfarist* terms: a Pareto improvement is good because everyone's welfare is improved;

forcing everyone to cooperate is a Pareto improvement and therefore coercion is warranted. If a coercive intervention could be avoided because an alternative arrangement is feasible, which would bring about the exact same level of public good without coercion, welfarism would be indifferent between the two. This specific evil of coercion is therefore not recognized at all by traditional welfare economics. And yet, as seen during the recent pandemic, its large-scale use by governments sparks debate, protests and outrage among many citizens. Is there nothing to object to opponents of lockdown measures deeming them liberticide?

The opposite problem is found in a completely different approach, namely the brand of libertarianism inspired by Nozick (1974). In this approach, coercive state intervention to force individuals to finance or produce public goods is always wrong, even if the resulting situation is vastly preferable to the *status quo* in terms of welfare. In the scheme sketched by Nozick, individuals are endowed with rights that impose constraints on actions that anyone else may legitimately perform[1]. The fact that, according to Nozick's libertarianism, individuals have the right to dispose of their bodies and possessions as they see fit imposes the obligation for everyone, including the government, to refrain from doing anything that might violate these rights, even if it means that vital public goods are not financed or produced. Rights define what individuals may do, at any point in time, and any situation that results from these actions on the part of individuals is just, provided that the initial situation was itself just. Legitimate actions thus 'preserve' the justice of the initial situation, as well as the freedom of individuals—which consists in not being prevented from doing anything they might want within the limits imposed by their rights. The use of coercion, on the other hand, implies *ipso facto* that the resulting situation is unjust, and that freedom is not preserved.

How can public goods be produced in a libertarian society? If contributions can only be voluntary, and the choice to contribute or not is decentralized, we may expect self-interested individuals not to contribute, as in a classic public good problem. The public good would therefore not be produced, without this creating any injustice or loss of freedom according to libertarians. Yet it would be wrong to conclude that libertarianism has nothing to say about public goods. A third party may well propose to every individual concerned by the public good a voluntary *assurance contract*

[1]Hence the expression 'rights as side-constraints' (Nozick 1974, 29).

(Schmidtz 1994, Tabarrok 1998) by which individuals commit *in advance* to contribute to the public good on the condition that all others do the same commitment, and receive compensation if they do not. This circumvents the public good problem: at equilibrium, everyone makes the commitment and the public good is produced without anyone's rights being violated.

One of the weaknesses of this kind of non-coercive solutions to the public good problem is that they may have a huge opportunity cost—the cost of not forcing people to contribute—, especially in catastrophic situations. In what I will call 'situations of urgency', it would take a long time to propose and implement the contract (because of the need to spread information about the contract and convince people to sign it), whereas the value (or the cost) of the public good may decrease (or increase) sharply over time. An immediate coercive intervention would thus be much more efficient. An exemplary case of this kind of situation is that of the start of a deadly epidemic in a locality: local residents may slow down or even stop the progression of the epidemic by carrying out certain actions that are very costly for them, such as isolating themselves for several weeks. The public good—which is the disappearance of the epidemic—would be considerably more difficult to produce once the epidemic spreads to the population: the number of people who need to be confined increases exponentially, etc.

Other examples of such situations may be:

- an imminent flooding requiring the immediate edification of a dam;
- a military invasion requiring to resort to compulsory mobilization of individuals;
- a global climate warming with very serious negative feedback loops requiring that everyone reduce their carbon emissions immediately.

Let us assume that we are in such a dramatic situation of urgency. The provision of a public good by setting up an assurance contract is no longer something that can be seriously considered. Either the public good is produced by compelling thousands individuals to contribute, or nothing is done, and individuals fail to coordinate to produce the public good. Nozick (1974) acknowledges that it may be desirable to suspend the obligation to respect individual rights in the event of a 'catastrophic moral horror'[2], but does not go on to specify what would make such violations acceptable.

[2]'The question of whether these side constraints are absolute, or whether they may be

We can go further by noting, with Sen (1992), that even if individuals facing such a public good problem *cannot* really coordinate to contribute and produce the public good, they *would* choose to do it if they could, even if it means that they would have to sacrifice some of their freedom. A coercive intervention could therefore, in a sense which needs to be specified, impose nothing on individuals that they would not, in certain circumstances of their choosing, do by themselves. But how can there be freedom where individuals have no *control* over their actions? To make sense of this idea, I will prolong Sen's efforts to define what he called 'indirect liberty' (Sen 1982) and integrate it into an 'extended' libertarian framework to derive the conclusion that coercive state interventions, and in particular lockdown measures, are not necessarily liberticide.

The chapter thus contributes to the debate about the merits of public or private provision of public goods by showing how the public provision of a public good *via* some form of coercion can still be compatible with freedom. It does so without resorting to traditional welfarist evaluation, but relying on what I call 'extended libertarianism', which is capable of justifying, under conditions that will be specified, a coercive intervention. This was already Serge-Christophe Kolm's goal when he attempted to define what he called a 'liberal social contract' in a somewhat forgotten book (Kolm 1985). A liberal social contract is a hypothetical contract to which individuals could have consented, to produce some specific results such as the provision of public goods. However, Kolm did not say much about the exact conditions that would make such a counterfactual consent valuable from the perspective of freedom, which this chapter intends to do.

In the first section, the chapter describes the structure of the problem of producing public goods that will be considered next, and how it may be done in a libertarian society. In the second section, the article introduces Sen's idea of 'indirect liberty', and his critique of the restriction of the meaning of freedom to what he calls 'freedom as control'. Someone's indirect liberty is preserved if, although they are not in control of the decision, they get what they would have chosen. The article develops Sen's idea by presenting the elements of an ethics of simulating choices, which is concerned with the permissibility, for a third party, to do certain things for individuals that they would choose to do themselves if they could. Finally, the third section applies

violated in order to avoid catastrophic moral horror, and if the latter, whet the resulting structure might look like, is one I hope largely to avoid' (1974, 30).

the elements of this ethics to the case of the public good problem to show how extended libertarianism, which allows for public intervention that simulates choices that individuals would have made, legitimizes the use of coercion to produce a public good in a situation of urgency.

2.1 Producing public goods in a libertarian society

The essential characteristic of a public good is that, once it has been produced, all individuals enjoy it in the same way, whether or not they have contributed to its production. The collective response to the kind of 'catastrophic moral horror' which I will consider here is a public good: everyone is saved from an epidemic, a war, a flooding, a brutal climate change in the exact same way. This collective response takes the form of coordinated rule-following behaviour expected from individuals: isolating oneself, accepting military conscription, building a dam, and respecting a set of individual quotas on carbon-intensive consumption. If enough individuals follow these rules, the 'catastrophic moral horror' is avoided and everything goes back to normal. But following them has a high cost for individuals: a loss of freedom, a loss of time and money, a risk of death, etc. If her effort is not *necessary* to build the collective response that averts the catastrophe, an individual will choose not to make them. Neither will she make these efforts if they are not *sufficient* to avert the catastrophe.

Therefore the essential feature of this collective response can be modelled as follows: (1) contributions from individuals are binary: they can either contribute, which is costly for them, or do nothing; (2) the public good is itself binary. More precisely, I will assume that there is a certain critical number of contributions such that we can be sure that the public good is produced when this number is reached; (3) each individual prefers the situation where the public good is produced to the situation where it is not but, everything else being equal, they prefer not to contribute. (4) the *status quo* situation is such that the public good is not produced and no contribution has been made[3].

In this setting, a 'public good problem' arises. Let us suppose that there

[3]This description is in line with Tabbarok (1998), whose assurance contract I will consider later.

are N individuals and that the contribution of at least $K \leq N$ individuals will be enough to produce the public good. In this setting, the *status quo* where the public good is not produced and no individuals contribute is a Nash equilibrium, because contributing is costly and will only make the person who pays it worse off given that no one else is contributing. We can thus expect the status quo to persist, and the 'moral horror' is not avoided. At the same time, any situation where more than K individuals contribute is a Pareto-improvement from the status quo, since individuals prefer the situation where the public good is produced to the situation where it is not, whether or not they contribute[4]. If welfare is the only thing that matters, a coercive state intervention that forces at least K individuals to contribute is justified because it leads to such a Pareto improvement: everyone is made better off.

The 'public good argument' (Schmidtz 1991) therefore applies there. According to this general argument, since in a public good problem the *status quo* equilibrium is expected to persist if individuals are allowed to choose whether to contribute or not, and that forcing individuals to contribute is a Pareto-improvement, it is necessary to force individuals to contribute in order to improve the situation of everyone. As coercion is *necessary* to reach this desirable outcome, it is also *justified*. The public good argument is supposed to give normative validation to coercive state interventions aiming at producing public goods. However, libertarian-minded economists and philosophers have devised solutions to the public good problem that disprove the public good argument—by showing that the premise that coercion is necessary is false. In particular, there exists now a whole class of 'assurance contracts' which shows how the public good problem can be addressed by a voluntary—'private' but coordinated—kind of provision. Assurance contracts enable people to coordinate to produce the public good, by making one's contribution conditional on the contribution of others, and by giving individuals incentives to sign the contract even if they expect it to fail. I will describe Tabbarok's (1996) particular solution to the problem, as it is perfectly suited to the public good model I consider.

The contract has two steps. In the first step, an entrepreneur offers the contract to every one of the N individuals, who is free to accept or reject

[4]However there is no clear case of optimality to be defined here if there does not exist a precise threshold such that the public good is produced if is reached and not produced if it is not. In my presentation, the number K is not such a threshold. Even if we can be sure that the epidemics disappear if ninety percent of the population self-isolates for two weeks, it does not mean that a smaller number would not be as effective.

it. We thus have two possibilities: either a number $X \leq K$ of individuals have entered into the contract, and the contract is said to have succeeded, or this is not the case and the contract is said to have failed. If the contract has succeeded, the X individuals who have entered into the contract are required to contribute, and if they do so the public good is produced. If the contract has failed, the X individuals who have entered into it are not required to contribute, but they receive a small payment—and of course, the public good is not produced. The only subgame perfect equilibria of the corresponding game are the situations where exactly $X = K$ individuals enter into the contract. As before, any such equilibrium is a Pareto improvement from the *status quo* situation. It makes everyone among the K contributing individuals necessary and sufficient to produce the public good. In other words, every contributing individual is *pivotal*.

The mechanism behind Tabarrok's 'dominance assurance contract' is intuitive:

- the situation where strictly less than K individuals sign the contract is not an equilibrium since any one of the others has an incentive to also sign it, either to receive the small payment or to make the contract succeed and produce the public good;
- the situation where strictly more than K individuals sign the contract is not an equilibrium either since any one of these individuals has an incentive to deviate and not sign it as they would not have to contribute but would still benefit from the production of the public good;
- when exactly K individuals sign the contract, any one of the others have no incentive to also sign it as they benefit from the production of the public good at no cost for them, whereas every one of the K contributing individuals, being pivotal, cannot deviate without making the contract fail and preventing the public good from being produced. We therefore have an equilibrium.

Assuming that transaction costs are not too high, such an equilibrium could easily be reached, which leads to the conclusion that coercion is not necessary: the public argument would fail to justify coercive state interventions in any situation where dominance assurance contracts could be implemented.

However, I am concerned here with what I call 'situations of urgency', which rule out the implementation of this non-coercive solution, because:

- collecting contributions can take a very long time, due to the procedures involved in drawing up the contract, reaching potential contributors and convincing them, etc.

- The value of the public good may decrease rapidly over time, or the public good may be more difficult to produce (more contributions required, or at a higher level, or the population concerned by the public good is larger) over time.

I will suppose that the opportunity cost of waiting for a dominance assurance contract to be implemented voluntarily is too high to be paid—we would be falling quickly into a 'catastrophic moral horror' if we simply wait. This particular context gives relevance to the public argument, as coercion appears necessary from a *moral* (and not a technical) point of view. For moral reasons, individuals cannot be left to coordinate voluntarily as they would do if a dominance assurance contract could be implemented. In such situations of urgency, the public good argument is—as seen during the recent pandemics—often formulated in terms of a trade-off between freedom and other values. Coercive safety measures are justified by insisting that the momentary or limited loss of freedom experienced is more than outweighed by the expected gain in terms of lives saved and lower pressure on hospital services. This does not address the concern that these measures are profoundly *liberticide*, which can lead some people to object to these interventions, even if they agree that preserving the *status quo* situation is not desirable and that they would change it if they could. But, as I will show, the case for a coercive state intervention need not be formulated in terms of such a trade-off between freedom and welfare or other values, which libertarians or freedom lovers may refuse. Such an intervention can be justified purely in terms of freedom.

Broadly speaking, there are two ways of assessing the freedom of individuals in an economic context. The first is based on the measurement of choice sets, or opportunity sets. According to this approach, each individual has various opportunities, which are things that she can bring about if she chooses to do so. The set of all these opportunities is her opportunity set. In principle, if we have a satisfactory metric for measuring these opportunity sets in terms of freedom[5], it is possible to compare each opportunity set with

[5]A literature has emerged in normative economics and social choice theory to complete this task, which I will call the 'freedom of choice literature'. See Barberà et al. (2004) for

any other, and thus to compare the opportunity sets that individuals have when the *status quo* situation is preserved with those that they would have in the situation where the public good is produced because of a coercive intervention. If we find that everyone has more freedom in the latter situation, we could conclude that the intervention is actually improving freedom globally. But carrying out this type of analysis is very difficult because:

- Some metric or rule must be chosen to measure each individual's opportunity set, but there is no consensus in the freedom of choice literature to favour one or another. Another difficulty is that of identifying all the opportunities accessible to an individual, which is not an easy task.

- a coercive intervention would close off certain opportunities at some point in time (such as seeing one's friends during an epidemic), and open up other opportunities later (living a life free of epidemic disease)—compared to maintaining the *status quo* situation. We would thus need to compare *sequences* of opportunity sets, which is more difficult than comparing opportunity sets. How would the trade-off between having fewer opportunities before the intervention, but (presumably) more after, be represented? To my knowledge, there exists no framework proposing to describe intertemporal freedom tradeoffs convincingly.

Another approach to freedom evaluation, inspired by Nozick (1974), would completely evade these difficult—if not intractable—questions. This approach is not based on comparisons and does not attempt to measure freedom. It remains agnostic on the question of whether a certain social or economic change increases or decreases the opportunities available to individuals. The evaluation is binary, in that it only asks whether or not a change preserves the freedom of every member of a society. In the Nozickian version of libertarianism, individuals have rights (to dispose freely of the things they own—including their own body—, by giving it away, exchanging it, etc.). These rights can be exercised in whatever manner that pleases individuals. individuals are free, in a negative manner, when they are not prevented by anyone to do whatever they want within the limits defined by their rights. If a social change happens in such a way that it does not violate anyone's rights, it can be said to preserve the freedom that everyone had before the

a survey.

change (if they were already free), since it does not prevent anyone from doing whatever they want within the bounds of their rights. Although Nozick does not insist much on that point, we can identify a freedom-preserving social change without having to specify which opportunities are opened to whom.

Compared to standard welfare economics evaluation practice, this approach has another particularity: because of its non-consequentialism, the focus of the evaluation is not on results (or 'alternatives'), such as a given allocation of goods in an exchange economy, but on social changes. In order to know whether a social change is freedom-preserving or not, the question is only whether such changes violate or respect individual rights. But in order to know whether a society is free, globally, as a result of this change, we would also need to know if it were free before it. If this was not the case, there would be nothing to preserve. According to Nozick's 'historical' conception of justice and freedom, a society is free if it started from a just initial attribution of rights (in particular, property rights over natural resources), and evolved through social changes that never violated any one of them. This makes this conception particularly demanding and inconvenient for evaluating whether a society is free or not, as past violations of rights would make virtually any society unfree. But we can still evaluate whether or not a change is freedom-preserving.

Since our goal is essentially comparative, there is no need to endorse Nozick's full historical conception. Provided that the situation before the social change was free enough[6], in a sense that does not need to be further determined, and that individual rights are well defined, it is possible to conclude that a social change is freedom-preserving, and that society is at least as free as it were before the change—just as the conclusion of a deductive reasoning is as true as its premises, provided that it logically follows from them. The comparative nature of the evaluation we need to conduct makes it admissible to get rid of the most controversial aspect of Nozick's conception while conserving its most appealing feature, which is its simplicity. If we can be reasonably sure that the society in which we find ourselves is free *to a*

[6]What would make a society free, initially? One possible answer would be that a society is free if it guarantees a certain number of fundamental social opportunities to each of its members, enabling them to lead their lives as they see fit—Sen (1995, 67) refers to Berlin's emphasis on the 'liberty to choose to live as he or they desire' (Berlin 1969/2002, 215)—and excluding the exploitative situations that Nozick's libertarianism would allow. Defining precisely these fundamental opportunities is obviously difficult.

certain degree, then a social change that respects everyone's right preserves that freedom, provided that we accept two assumptions essential to Nozick's approach:

- A natural change (such as an earthquake, flooding, etc.) that reduces individual opportunities cannot affect freedom, since such a change does not violate anyone's right, at least directly. A natural disaster that destroys a country's infrastructure and economy undoubtedly makes people's lives miserable but does not affect freedom.

- When individuals choose actions that have the consequence of reducing their own opportunities (in a way that is compatible with their rights) or of reducing others' opportunities, society is still as free as it was before. What matters is that individuals are able to exercise their rights, and not the consequences of how they exercise them: this is the product of Nozick's non-consequentialism. Thus, the appeal of this conception depends largely on the way rights have been defined.

While this approach, because of its crude binary character, cannot constitute a viable alternative to the evaluation practices of standard welfare economics, it can complement it because it allows us to make judgements about freedom that are simply based on the information that we have about rights violations. A social change is not freedom-preserving if it violates someone's rights. It is freedom-preserving if it does not violate anyone's right. An example of such social change is the design and implementation of a dominance assurance contract, as defined earlier. A third party (the state or an independent entrepreneur) may propose the contract to anyone likely to sign it and then ensure compliance with the terms of the contract. Even if such an enforcement operation is likely to meet some resistance from people who had initially accepted but changed their minds, it does not violate any libertarian right, since individuals have the right to enter into contracts by which they bind themselves by promising to do or deliver something in the future.

The consent that makes the contract valid (according to libertarians) encapsulates a different kind of information than the purchase of a good, for example. When deciding whether to consent to signing such a contract, an individual may have in mind:

1. the *final result* that she gets for herself.

2. the *process* by which change occurs for the individual. In particular, if the final results that she gets for herself are not enough to offset the loss of freedom that she would incur by binding herself (in our case, by promising to contribute to the public good at a later date), she may refuse the contract even if she values its final result.

3. the final results that *other* individuals get. In particular, if the *distribution* of resources or burdens (in our case, the burden of contributing to the public good) that the individual expects the implementation of the contract to generate violates her sense of justice, she may refuse it even if she values its final result for her.

In consenting to such a contract, an individual determines that, regarding the three previous aspects, combined together as a whole, signing the contract is better than doing nothing (and maintaining the *status quo*). The information that this consent would reveal is much richer than what the preferences defined over final results that are the basic inputs of the evaluation in standard welfare economics encapsulate[7], as, crucially, it also says something about the *trade-offs* that individuals are willing to make, in terms of welfare and freedom. The person who consents to bind herself and lose some significant opportunities in the hope of attaining a better situation is making such a trade-off, which indicates what is an acceptable compensation for the loss of her opportunities. By contrast, limiting the inputs of the evaluation to preferences defined over final results overlooks the fact that a change may be considered unacceptable because it involves losing too much freedom, even though the individual is better off as a result of this change[8].

In Nozick's own version of individual rights, a coercive state intervention always violates the rights of individuals, even if it does not reduce individual freedom more than what they would be ready to accept in the context of a voluntary transaction. If we want to determine how such an intervention may be acceptable to avoid the 'catastrophic moral horror' that a situation of

[7]and that, for instance, a series of purchases of private goods would reveal.

[8]In a different context, Fleurbaey (2006, 303-304) argues that a capability metrics based on set evaluation would lose sight of achievements by focusing exclusively on opportunities. He recommends focusing instead on what Sen called 'refined functionings', which consists of the pair (capability set, achieved functioning vector). The kind of information I have in mind here would consist of a triple (former opportunity set, new opportunity set, final outcome), encapsulating the trade-off that individuals would make between losing opportunities and attaining a valuable outcome.

urgency may produce, we will need to change the definition of rights that is the core of Nozick's conception so as to amend his libertarianism. I propose to call this amended version 'extended libertarianism'. I will define it in the next section.

2.2 An ethics of simulated choices

What individuals control, in a public good problem, is essentially their choice of whether or not to contribute to the public good. What is beyond their control (except in very special circumstances) is the production of the public good. And yet, producing the public good is something individuals would choose to do if they could really do it. In this sense, it matters for freedom, provided that there can be freedom where there is no control. This point was made by Amartya Sen:

> The freedom to live in an epidemic-free atmosphere may be important for us, and given the choice, we would choose to achieve that. But the controls of general epidemic preventing may not be in our hands—it may require national and possibly even international policies. If we do not have control over the process of elimination of epidemics, there is no more to be said, as far as our 'freedom as control' is concerned, in this field. But in a broader sense, the issue of freedom is still there. A public policy that eliminates epidemics is enhancing our freedom to lead the life—unbattered by epidemics—that we *would choose* to lead (Sen 1995, 65)

Unfortunately, Sen does not describe in greater detail this counterfactual choice, which he considers relevant to assessing an individual's freedom. It is the reference to this counterfactual choice that allows Sen to conclude that the end of an epidemic improves individual freedom. The choice of stopping the epidemic is not a choice that someone can make on his own under normal circumstances, but it is a choice that one would want to make, and would make, in circumstances where one would have this opportunity. To elucidate the nature of the counterfactual choice that Sen alludes to, we need to identify those circumstances, and the exact trade-off that people would be willing to make to be able to live in an epidemic-free atmosphere.

As we shall see, the fact that these circumstances do not arise is a product of the public good problem, which prevents individuals from cooperating to produce the public good (especially if we are in a situation of urgency). The public good can be immensely valuable for individuals, and that may justify a coercive state intervention aiming at coordinating individuals' efforts to produce it. And if it can be justified in terms of freedom, the notion of freedom we need would not be tied to that of control, actual choice and opportunity. My claim is that coercion can preserve freedom, provided that the restrictions that are imposed on individuals are exactly the ones that they would impose on themselves if they were sufficient to produce the public good (that is, if the production of the public good were under their control). To accomplish this task, I will need to broaden the concept of freedom which is implicit in Nozick's version of libertarianism.

The libertarian approach is tied to a notion of 'liberty as control', as Sen puts it[9]. According to this conception, 'a person's liberty is related to the extent of the control that he or she has over decisions in certain specified spheres' (Sen 1982, 207). Contracts and other voluntary transactions are exercises of this freedom as control, because if someone has agreed to restrict their freedom in the future to get something in return (as is often the case with contracts), they have consented to everything they are bound to do. Coercion, on the other hand, take the levers of control out of individuals' hand. If we stick to this idea of freedom as control, it will be impossible to understand how building a collective capacity to put an end to an epidemic can enhance—or, as I will argue, preserve—individual freedom.

Sen's key argument against the notion of liberty as control is that there are many situations where individuals are not in control, and yet freedom is at issue. Consider Sen's example of someone who, after an accident, is left bleeding and unconscious, in need of a blood transfusion[10]. A conception of freedom as control does not give us any guidance about what should be done to her out of concern for her freedom since the unconscious individual is no longer in a position to exercise her right to receive or refuse the transfusion. Yet someone who knows the person reasonably well could tell us if she has, for example, religious objections to receiving the transfusion. By considering the choices the person would have made had she been conscious, we extend

[9]Sen has in view this particular passage from Nozick: 'Individual rights are co-possible; each person may exercise his rights as he chooses. the exercise of these rights fixes some features of the world' (Nozick 1974, 166).

[10]See Sen (2002, 396).

that person's capacity to choose—and her freedom—to this situation; we can simulate her choice, just as if she were there. However, such a judgment in terms of freedom can hardly be justified on the basis of the notion of freedom as control. There is nothing that such a conception could tell us about what to do in this circumstance. We therefore need to add what Sen calls 'indirect liberty' or freedom.

According to Sen, we can say that someone's indirect freedom is better 'served' (Sen 2002, 396) in the case where she receives the transfusion, than in the case where she does not, provided that she would have chosen to receive it. To adapt Sen's perspective to Nozick's framework[11], I will say that receiving the transfusion 'preserves' the indirect freedom of someone who would have wanted it. Respecting someone's counterfactual choice in his absence simulates the exercise of a right that the individual would have chosen to exercise had he been present.

From a classic libertarian perspective, deciding for someone else seems to be permissible only if an act of *delegation* has been made. Certain decisions can be delegated to a proxy agent (as in the case of proxy voting), who is allowed to simulate the decisions that a person would have taken if she were able to do it[12]. In the blood transfusion example however, something crucial is missing: the decision to delegate has not been made—if the person had

[11]The distinction between 'liberty as control' and 'indirect liberty' was made by Sen in response to an objection raised by Nozick about Sen's 'Paretian liberal' theorem. In essence, Nozick objected to Sen's definitions of rights in a social choice theory framework that rights 'fixes some feature of the world' prior to the application of a social choice procedure. Rights put constraints on the possible outcomes of the procedure. Sen responded that the social choice theory perspective can still be useful especially when 'liberty' is not conflated with control. The goal to integrate this 'indirect liberty' into the libertarian framework may thus appear surprising. But the value of Sen's argument can be acknowledged by libertarians since it brings to the fore the impossibility for advocates of 'liberty as control' to say anything relevant about fairly common situations (such as that of the unconscious person). This shows the limitations of the libertarian perspective on freedom.

[12]This process of simulation, and the reference to some act of delegation, is also present in Sen's analysis, who uses the example of a proof-reader: 'The proof-reader will be doing what I would, counterfactually, have done if I were to correct all the proofs myself with eyes as efficient as that of the proof-reader' (Sen 1995, 64). The proof-reader simulates the decision that I would have made. Even if Sen does not say it, he is normally allowed to do it because I (the author) have agreed to let him do that. But Sen seems to imply that the simple fact that some choices are effectively simulated (with or without consent) is sufficient to conclude that our freedom is enhanced. I will propose here a more ethically demanding ethics of simulated choices.

When delegating a decision to someone else, I	Blood transfusion example:
(1) choose to delegate (and some agent to act on my behalf)	I authorize physicians...
(2) choose a set of circumstances in which this person is allowed to choose	... in matters relevant to blood transfusion and if I am unconscious...
(3) choose what she will choose in these circumstances.	to perform the transfusion

Table 2.1: Indirect liberty and delegation

already stated that she wanted to receive a blood transfusion, we would still be in the realm of freedom as control. But suppose that we could be reasonably sure that the person would have made this act of delegation—letting physicians take care of her body and perform the transfusion. What we need to make sure is that every aspect of this act of delegation would have been consented by the person. The decision to delegate can be broken down into different choices, that would need to be simulated. They feature in the left column of table 2.1.

The notion of indirect liberty or freedom[13] involves that it is sometimes legitimate for a third party (here, the physicians) to intervene in someone's life, even though she has not authorized them by an explicit act of delegation to decide her place. I will call such an intervention which replicates exactly the choices that someone would have made 'simulated choice'. Simulating choice is relevant only if the person is really not in a position to make those choices and an act of delegation could not have taken place—the first choice described in the above table cannot be made. This reflects the idea, also shared by Sen, that freedom as control has some priority over indirect freedom.

- *Condition 1.* An intervention simulating choices is permissible only if the person concerned did not already voluntarily delegate this choice to someone else, and was not in a position to do it.

The second condition is not mentioned by Sen, but it is crucial to make an intervention simulating choices as closely as possible as an act of delegation.

[13]I use the two terms interchangeably.

Individuals choose to delegate some decision tasks to others because they would get bad or worse results without these others intervening to simulate their choices. But this delegation decision comes with many strings attached: a delegation is not an abdication of someone's will, but its expansion. An act of delegation would specify a future set of circumstances in which the proxy agent is allowed to choose, which determines the range of decisions that can be taken by this agent—this is the second choice described in the above table. One can represent this choice as a choice between opportunity sets: someone trades off the opportunity set he would have without delegating for the opportunity set he would have if he were the proxy agent. Someone who cannot cast a vote at election time because he must attend a funeral elsewhere can delegate to a proxy agent the task of voting for his favourite candidate. Without proxy voting, this person can only choose between casting a vote or attending the funeral. With proxy voting, he can do both (and is better off as a result). He has exchanged a less valuable opportunity for a better one. In simulating someone's choice of a future set of circumstances in which the proxy agent would choose on his behalf, we must make sure that the individual would be ready to trade off the opportunity set that he would have without the intervention for some opportunity set that he could have (through the intermediary of the proxy agent) if it were possible to delegate[14].

- *Condition 2.* An intervention simulating choices is permissible only if the person concerned would choose to exchange the opportunity set he would have without the intervention for some opportunity set (call it O), accessible only to a proxy agent, that contains the alternative that the intervention implements.

The intervention should therefore simulate the decision to trade opportunity sets which is implied in the act of delegating. It should also, of course, simulate the choice that the individual would want the proxy agent to make. This is the third condition:

- *Condition 3.* An intervention simulating choices is permissible only if the person would choose the alternative (call it a) that the intervention implements in the opportunity set O.

[14]This discussion supposes that individuals' preferences over opportunity sets are stable over time: at any point in time, an agent rank two sets in the exact same way. Therefore, whatever may be the circumstances of the decision of delegating, the choice of a set would be the same.

The second condition is essential to define extended libertarianism and compare it to other approaches. Some economists and philosophers have already explored the issues I am raising here (e.g. Duflo 2012, Sunstein 2019). Esther Duflo, in particular, defends a form of paternalism based on the idea that it is desirable to avoid imposing certain choices on individuals that unnecessarily complicate the decision-making process or are time-consuming. For example, for Duflo, the fact that individuals in poor countries have access to both non-potable water and potable water (by boiling the former, for example), whereas individuals in rich countries only have access to potable water (as they would have to go through some complications to get some non-potable water) does not show that the former are freer, or better off, than the latter. The undesirability of non-potable water—the fact that, in Sen's terms, nobody *would* choose to drink non-potable water—, would justify removing this option and switching to the situation where everyone can only drink potable water.

From Duflo's point of view, then, it would not be illegitimate to restrict the choice of options through coercive state intervention, provided that we eliminate only parasitic options, which only make the decision more complex because they are undesirable—in the sense that nobody would choose them under reasonable circumstances. This point of view is, of course, at odds with the classical libertarian approach, since the elimination of undesirable options would violate the rights of individuals to retain them if they correspond to a rightful exercise of their rights. It is true that, in a libertarian society, individuals could freely agree to give up these parasitic options. They could thus 'choose not to choose' (as Sunstein 2015 puts it) and make arrangements not to have these options or to delegate to another agent the task of doing these choices for them. People who hire life coaches, personal assistants or rely on family members to make decisions on their behalf do exactly that.

But the possibility to delegate, which a libertarian framework offers, is probably not enough, in Duflo's view, as this decision, or more generally the decision to 'choose not to choose' is itself costly as it requires time and energy, particularly from poor people who have a limited psychological 'bandwidth' because of poverty[15], but may be those who need it the most. A coercive intervention that restricts individual choices may therefore be desirable to restore what Sunstein (2019) calls the 'navigability' of individual choices.

[15]'Bandwidth captures the brain's ability to perform the basic functions that underlie higher-order behavior and decision' Schilbach et al. (2016).

The intervention would implement the alternative that individuals would choose, simplifying the decision process and making their lives easier. But from the perspective of extended libertarianism that I just defined, it is not enough to point at evidence that some options are parasitic, or make the choice process too complex or time-consuming, to make such an intervention permissible. It is not enough that people get what they *would* choose as a result of the intervention; for there is no assurance that individuals would be willing to trade off the restriction of their choices for the outcome of the intervention. However, if they would accept to delegate these difficult choices to someone else, but could not do it because they are trapped in poverty or do not have enough 'bandwidth', we can be sure that we did not impose on individuals more constraints than they would impose on themselves.

The essential difference between Duflo's paternalism and extended libertarianism's justification of a coercive intervention is therefore that the latter requires, in addition to the fact that individuals cannot really avoid some difficult choices by making other people make decisions on their behalf, that *they would if they could*. In this sense, extended libertarianism is more ethically demanding and requires more information on individuals' counterfactual choices. It is the price to pay to be able to preserve individual freedom, in a meaningful sense.

Let us now return to the case of public goods. An intervention simulating choice would have to be based on a delegating decision which would involve multiple individuals, as an isolated individual cannot produce the public good by herself. In a libertarian society, individuals may accept Tabbarok's dominance assurance contract, which gives everyone an incentive to contribute to the public good. Acceptance of this contract can be seen as a form of delegation: by promising to contribute if the contract succeeds, the individual allows the executor of the contract to make him contribute (or not) in the situation where the dominance contract succeeds—that is, in the situation where enough individuals have accepted the contract to make it work. It is no longer, in this case, up to the individual to choose to contribute or not, and at the same time, the choices are not the same as before, since now contributing leads to the production of the public good. The situation is therefore formally similar to that of a delegation, as I have described it. If individuals would agree to sign a dominance assurance contract, they would satisfy the two last conditions that I defined earlier.

However, the fact that the burden of contributing must be shared among different individuals adds another layer of complexity. An additional con-

dition must therefore be added: *who* is asked to contribute may indeed be of importance for the person who considers signing the contract. Suppose that all individuals would accept the contract, but the executor would only ask that the poor, or some particularly disadvantaged part of the population, to contribute. As some would find this particularly unfair, they would not consent to sign such a contract and would prefer that the public good should not be produced at the cost of such injustice. We must therefore add a fourth condition for a coercive intervention based on simulated choice to be permissible:

- *Condition 4.* An intervention simulating choices is permissible only if the person who would choose a in opportunity set O knows the distribution of benefits and burdens that such a choice entails.

This knowledge of the final result of the person's choice for others guarantees that she actually consents to the distributive consequences of this choice. We must be sure that all the relevant trade-offs (between one's freedom, or other people's result, and one's final result for oneself) have been done by individuals. Otherwise, we would ignore a relevant source of concern for individuals, which libertarian freedom addresses (since individuals can refuse to engage in a social change that has distributive consequences that they find unfair), and that extended libertarianism should also acknowledge.

I claim that these four previous necessary conditions are jointly sufficient to make a coercive state intervention permissible. This would allow for the existence of a more-than-minimal state. Rights are, in Nozick's framework, materialized by a set of constraints imposed on the actions of individuals. They are all the constraints that an individual' right imposes on the action of others, as required by the doctrine of 'rights as side constraints'. For a coercive state intervention to be possible in such a framework, it must be that the state has some rights of its own, which are not the result of past transfers of rights from individuals and imposes some constraints on the actions of individuals—such as the obligation to contribute to the public good in a situation of urgency. But such a right arises *as if* it was the result of a transfer of rights from individuals. In that respect, extended libertarianism remains distinctively libertarian.

I will now reformulate the four conditions above so that they can be applied to a public good model, which will be done in the next section. Let us call S_1 the *status quo* situation, S_2 the situation post-intervention, $C_i(S)$

the opportunity set that individual i has in some situation S. The rights of individuals should be defined such that they would not be infringed by an intervention which would take us from S_1 to S_2 and be such that:

1. S_1 is a situation of urgency.

2. every individual i concerned by the intervention would accept to exchange $C_i(S_1)$ for $C_i(S_1')$, where S_1' is a *hypothetical* situation such that the opportunity sets $C_i(S_1')$ of every i are mutually compatible—provided that they would get what they *would* choose in $C_i(S_1')$.

3. the choice that every i would make in $C_i(S_1')$ produces situation S_2 and i knows it.

The first condition ensures that the coercive intervention was unavoidable: individuals could not set up a dominance assurance contract to produce the desired result. The second condition considers a hypothetical situation S_1', which provides every individual with opportunities that they would not have in a *status quo* situation. As was emphasized earlier, we delegate a choice to someone else is because the set of choices that the proxy agent can make (in our name) is better, from our point of view, than the set of choices that we would have without delegating. The second condition reflects exactly that: individuals would be ready to trade their opportunities in the *status quo* situation for the opportunities they have in the hypothetical situation—which is hypothetical because, in a situation of urgency, these opportunities would never be directly available to them. The fact that opportunity sets are mutually compatible ensures that the intervention simulates choices which could really have taken place. Finally, the fact that every individual knows that the opportunity selected in the hypothetical opportunity set will lead to the post-intervention situation (condition 3) ensures that consent has been given to let the distributive consequences of the intervention happen.

To summarize the argument: extended libertarianism states that it is legitimate to use coercion only if the result of a coercive intervention simulates the series of choices an individual would make if he were in a position to delegate a decision he would want to make, and would actually make that delegation. Libertarianism considers such delegation legitimate, but it is sometimes made impossible or very costly by circumstances. It is therefore desirable to amend libertarianism, to allow for a coercive intervention which

would not restrict freedom more than what individuals would accept themselves if they could make this act of delegation. This type of intervention does not preserve freedom as control (which is characteristic of libertarianism), since coercion violates the rights of individuals, defined in a standard libertarian way. But it does preserve indirect freedom, in the sense that individuals would have made the relevant trade-offs between the final outcome and a certain loss of freedom that the intervention simulates. But because of the priority of freedom as control over indirect freedom, extended libertarianism, which gives an important role to the state, only applies when delegating is impossible even if it is desirable, in particular when we are in a situation of urgency.

2.3 Application

How do the examples of a collective response to disasters which were presented in the introduction—building a dam to avoid imminent flooding, creating a conscript army to fight an invading military power, setting up carbon quotas to avoid fast global warming, setting up a lockdown to prevent the spread of an epidemic—relate to the model I proposed in the first section? Among the population that is concerned by the public good and can contribute, the contribution can be seen as binary, and the public good is produced when a certain amount of contributions are made. In every case, it is possible, without needing too much information, to fix a threshold such that we can be reasonably sure that if it is reached, the public good is produced. In each of these situations, waiting for too long has a huge opportunity cost because of the decrease (increase) of the value (cost) of the public good with time. We thus are in situations of urgency, which would, according to the public good argument, justify a coercive intervention. But how could it preserve freedom?

We need to check if the three previous conditions defined in the last section are verified. Since the examples are chosen to exemplify situations of urgency, the first condition is satisfied. To see if it is also the case for the other conditions, some assumptions about individual preferences need to be made. I will assume that, as in the structure of the public good game described in the first section, all individuals taken from the relevant population would prefer the situation in which the public good is produced, whether or not they have contributed to it. However, they prefer to contribute only when they

are *pivotal* in producing the public good—because they would pay the cost associated with their contribution for nothing if there are not enough people ready to contribute to the public good, or if there are already too much. It is precisely this structure of preferences that makes a collective response difficult to organize, but also possible to reach (otherwise there would be no hope of preserving freedom, as it was defined in the last section). The fact that a majority of governments of democratic advanced economies have chosen to impose strict lockdowns at the start of the recent COVID epidemic suggests that public officials believed at that time that key features of the public good model were relevant.

Indeed, as the *status quo* is a Nash equilibrium, we can expect it to persist in the absence of external intervention. The intervention to be evaluated consists of forcing a sufficient number of individuals to contribute, equal to the critical threshold beyond which we can be assured that the public good is produced. In the case of lockdowns, on which I will focus now, the public intervention sees to it that the major part of the population is forced to isolate themselves for a few weeks. At the *status quo* S_1, all individuals would prefer that the epidemic is ended, whether they self-isolate or not[16]. I will make the additional assumption—which is not part of the traditional public good game—that individuals would accept to sign a dominance assurance contract, as described in the first section, if they could. Since a dominance assurance contract provides them with the opportunity to be pivotal in producing the public good, we could expect individuals who value highly the public good and have the kind of preferences described in the previous paragraph to accept it. But in the 'extended libertarian' framework which I propose, one would also need to make sure that the cost in terms of loss of control that the contract involves is accepted.

At the *status quo* S_1, individuals have the opportunity to self-isolate or not, but doing so independently from others will not be sufficient to end the epidemics, as it is expected that other individuals at the *status quo* will not contribute. I will now describe a hypothetical situation S_1' which corresponds to the situation where a dominance assurance contract is proposed to a relevant number of individuals and where it is expected that, with effective

[16]Obviously, in a diverse society, preferences may differ and some people would never accept to self-isolate. But if such individuals are a small minority, as I suppose they are, we can restrict ourselves to the consideration of the majority who would self-isolate conditionally—because forcing people to do what they would never accept cannot preserve their freedom.

coordination, the contract will be successful. Such a dominance assurance contract would involve the obligation to self-isolate for several weeks for those K individuals, who, at equilibrium, accept it. The expected success of the contract involves that every individual accepting it is pivotal in ending the epidemics—which means that by being among the K individuals accepting the contract, one effectively has control over the end of the epidemics, since the contract fails if they withdraw from the contract (the *status quo* persists) and it succeeds if they sign it. In this hypothetical situation S_1', the associated opportunity set $C(S_1')$ contains the option to self-isolate and end the epidemics and the option to not self-isolate and return to the *status quo*. Compare it with $C(S_1)$, the opportunity set associated with the *status quo* situation: it contains the option to self-isolate without ending the epidemics, and to not self-isolate with the same result. In light of the previous assumptions, individuals would find it better to face the opportunities they have in the hypothetical situation S_1'. The fact that they would accept the assurance contract shows that they would prefer being in this hypothetical position where they are pivotal in ending the epidemic rather than maintaining the *status quo*. This could be rationalized by saying that $C(S_1')$ gives a higher indirect utility to the agents, or that in terms of freedom of choice $C(S_1')$ dominates $C(S_1)$ because the option to self-isolate and end the epidemics dominates the option to not self-isolate and maintain the *status quo*, and the option to not self-isolate and maintain the *status quo* dominates the option to self-isolate and maintain the *status quo*. From an informational point of view, what is needed is just that individuals would effectively exchange $C(S_1')$ for $C(S_1)$, which is guaranteed by the fact that individuals would accept a dominance assurance contract.

We also need to check if individuals would choose to self-isolate in the hypothetical situation, so as to produce the final situation S_2 where the epidemics ended whereas a number of individuals exactly equal to the critical threshold have also contributed. This is indeed the case, if individuals have the kind of preferences that a public good game assigns to them: producing the public good is all that matters if they can really control its production, which is the case here. Condition 2 is therefore satisfied. A number of individuals equal to the threshold K would accept to trade the opportunities they have in the *status quo* for those they have in the hypothetical situation, and then would choose to self-isolate in the latter situation, thereby ending the epidemics for all. A public coercive intervention would simply contract these two stages into one by imposing on these individuals to self-isolate, as

they would have preferred if we were not in a situation of urgency and an assurance contract was proposed to them. This produces directly the situation S_2. We know that individuals would willingly go through these stages, as the trade-off involved between freedom and the final result is exactly the one that they would make in accepting the dominance assurance contract.

A final important point to discuss is whether or not S_2 is seen as just, which is the question of whether condition 3 is satisfied or not. As we have seen, individuals may refuse a dominance assurance contract because it leads to unjust results, for example, because those who are asked to contribute are otherwise more disadvantaged than others (imagine that only ordinary people are asked to self-isolate, while rich people or politicians are allowed to party as hard as they want). A significant number of individuals would probably not have accepted the contract in these conditions. This implies that a coercive intervention could not preserve freedom, as it would force individuals to make a particular trade-off that they are not willing to make, and would not really simulate their choices. In practice, it may be difficult to make sure that people would not object to the distributive consequences of a particular coercive state intervention, but this condition is necessary for the intervention to be freedom-preserving under extended libertarianism[17].

Conclusion

To conclude, I will consider what could be replied, in light of the previous discussion, to someone who complains that a lockdown is liberticide. We could ask her the following questions: (1) would you like to be in a position where you could end the epidemic just by self-isolating? If your answer is yes, then in proposing you an assurance contract, the government would do exactly that. (2) if you could end the epidemics just by self-isolating, would you do it? If your answer is yes, then by forcing you to self-isolate, the government does exactly what you would have done to yourself if you had accepted the contract that you would have wanted to sign. In sum, the government, in a situation of urgency, imposes on individuals nothing more than what they would impose on themselves in a favourable situation that they would have chosen themselves. Under these conditions, a lockdown is

[17] A simple way to bypass the issue of unfairness would be to require that everyone who can contribute does it, which would mean that we fix $K = N$. This would change nothing to the logic behind the design of the intervention, but would be much less efficient.

not liberticide. It preserves the freedom of everyone.

Chapter 3

Paternalism for Rational Agents

In the context of strategic interactions, individuals sometimes find themselves better off when they have fewer options. This mechanism is known under the name of 'strategic commitment', as it usually is the individuals who 'commit' themselves to follow a certain course of action by restricting their options; but that is not necessary. I explain how a paternalistic intervention may be conceived where it is a third party who restricts rational individuals' choices to improve their welfare. This kind of intervention, which I call 'strategic paternalism', contradicts the narrative according to which welfare economics is incompatible with paternalism because it assumes individual rationality, which would make paternalism irrelevant. To prove this point, I show why and in what sense this 'strategic paternalism' deserves to be called that way.

Introduction

In recent years, a wave of articles and books written by behavioural economists and philosophers have advocated a paternalistic approach to public policy, based on the idea that individuals are often not capable of making decisions that are best for improving their welfare. *Libertarian paternalism* (Sunstein and Thaler 2003, 2008) has caught the attention by proposing that it is possible and desirable to influence people who are not fully rational (to 'nudge' them) without altering their freedom of choice. *Asymmetric paternalism* (Camerer et al. 2003) promotes public interventions that will affect

only those who are not behaving rationally, while not harming those who do. Advocates of libertarian or asymmetric paternalism see it as a form of 'soft' paternalism, which keeps individuals in control of the decisions they make. But some harder forms of paternalism have also been defended on the same grounds, notably by Conly (2013), who defends a new kind of 'coercive paternalism' warranted by individual reasoning failures. These proposals have generated a huge literature discussing the relevance of libertarian paternalism (Hausman and Welch 2010, Cozic and Mongin 2018), its moral permissibility (Grüne-Yanoff 2012), or worrying about the public policy trend that such a 'paternalistic turn' is setting (Sugden 2018a, Saint Paul 2011, Rizzo and Whitman 2020).

Advocacy of paternalism is not something particularly new in philosophy, but it is a novelty in economics. This paternalism differs much from traditional paternalism, which is associated with substantial religious or moral judgements and wide-ranging bans such as those experienced in the United States during the Prohibition era. Rizzo and Whitman (2020), in particular, contrast 'old' and 'new' paternalism, which they also call 'behavioural paternalism'[1]. What distinguishes the latter is that it turns to behavioural sciences to justify the interventions it advocates. We all are, if we follow Camerer et al. (2003), acting like idiots who need protection against their own reasoning failures[2]. A second feature of this 'new' paternalism is that the intervention should enable individuals to get what they want or would want for themselves[3], better than they could if they were left to choose by themselves without being influenced or coerced. behavioural paternalists do not try to impose their own goals on others, but they also do not take the goals expressed by individuals as given, since they may be distorted by biases or mistakes. The 'true' goals of individuals have to be inferred or reconstructed. In any case, this new approach is a challenge to the traditional principle of consumer sovereignty, which is central to welfare economics[4].

[1] 'We call the use of behavioural economics to justify paternalistic interventions "behavioural paternalism"'. (Rizzo and Whitman 2020, 16)

[2] 'behavioural economics extends the paternalistically category of idiots to include most people at predictable times' (Camerer et al. 2003, 1218).

[3] 'as judged by themselves', according to Thaler and Sunstein (2008, 5).

[4] The literature devoted to amending traditional welfare economics to take into account the lessons of behavioural economics is referred to as behavioural welfare economics. It must be distinguished from behavioural paternalism as some approaches described in this literature are not necessarily paternalistic. See Thoma (2021b) on the subject.

But how exactly is this a turning point for welfare economics? As Daniel Hausman (2021), among others, pointed out[5], welfare economics traditionally identifies individual welfare with the satisfaction of stable, context-independent preferences (a position often called *welfarism*), and assumes that individuals always choose what best satisfies their preferences. It would follow from this assumption of rationality that individuals' choices are always optimal. A paternalistic intervention that would restrict or influence these choices would only make things worse. According to this line of reasoning, paternalism has been excluded from traditional welfare economics *a priori* simply because agents are assumed to be rational in this sense. This sets the narrative that behavioural economics, having overturned this state of affairs—by pointing out that individuals do not choose what is best for them—, has thereby opened the gate to paternalism. In the following, I will challenge this narrative by arguing that conceptualizations from classical game theory have already made it possible to justify certain paternalistic interventions, even if this point was not explicitly recognised.

Haybron and Alexandrova have argued that in spite of what they call the 'normative minimalism' of traditional welfare economics—the attempt 'to keep value commitments to a minimum' (Haybron and Alexandrova 2013, 159) by using a preference satisfaction criterion—it is indulging in what they call 'inattentive paternalism' when it comes to applying cost-benefit analysis. By relying only on revealed preferences as an indicator of people's interests, economists fail to defer to individuals' judgements about their broader interests, neglecting the value commitments that are not revealed in their choices. But Haybron and Alexandrova note that 'minimalists' explicit intentions are non-paternalistic' (ibid., 167).

I claim that, on the contrary, economists' intentions may well be explicitly paternalistic if they draw the consequences from the lessons of Thomas Schelling (1960/1980) and his subsequent heritage in game theory—which belongs to the standard, 'minimalist', core of economics. The 'strategic paternalism' that I will describe, has not yet, to my knowledge, been explicitly sketched out. It is an obvious but insufficiently recognised consequence of the analysis of strategic interaction that rational agents are sometimes unable to make the best of the situation in which they find themselves. A 'strategic' intervention designed to help them better achieve their goals by interfering with their choices is in many ways similar to the interventions advocated by

[5]see also Hausman (2018), Saint-Paul (2011), Rizzo and Whitman (2020).

behavioural paternalism. But such an intervention may be justified even if individuals are rational, unlike interventions based on behavioural paternalism, such as nudges, debiasing, or bans of addictive substances.

The rest of the article will proceed as follows. It can be shown that something is missing in the logic of Hausman's argument that the welfare economics framework excludes paternalism, and I will lay this out in the first section of this article. The new behavioural paternalism which springs from abandoning the assumption of rationality is a variety of 'means paternalism', which I will describe in the second section. The distinction between means and ends offers a good way to understand how, in a strategic setting, individuals might fail to get what they want because of what they are. The exact mechanism by which such things occur will be detailed in the third section. The idea of 'strategic advantage' introduced by Thomas Schelling (1960/1980), can be used to make the case for a paternalistic intervention whose target is rational agents. The fourth section addresses the concern that such intervention is not needed because rational individuals could be provided with commitment device instead, which they could use to the same effect. I explain why sometimes rational individuals are not in position to use a commitment device to their advantage, which makes an intervention necessary. Finally, I will give in the conclusive section a general definition of what I call 'strategic paternalism' and show how it relates to behavioural paternalism.

3.1 The anti-paternalist argument

Why would traditional welfare economics be incompatible with paternalism? As Hausman puts it concisely:

> One advantage of the conventional view linking welfare to preference satisfaction is that questions about paternalism cannot arise. If what individuals choose is best for them, then is it impossible to make them better off by overruling their choices. (Hausman 2021, 19)

A certain definition of paternalism is implicit here. A paternalistic intervention is intended to overrule people's choices to make them better off. And this would fail, since overruling people's choices cannot make them better off. The crux of the argument, it would seem, is the rationality that is ascribed to

the economic agent: she always chooses what is best for her. More precisely, she is endowed with stable, context-independent preferences defined over her options[6], and always chooses the option that she prefers most. And according to the preference-satisfaction criterion, what she prefers is also what is best for her. Paternalism is therefore bad, or at least it is useless. But if we drop this assumption of rationality, we make room for something like behavioural paternalism: if individuals do not always choose what is best for them because of reasoning failures, biases, or problems of self-control, overruling these choices may make them better off.

But this argument is too quick to conclude. Three points of criticism will be raised. First, behavioural paternalists like Sunstein (2014) have insisted that terms like 'overruling' or 'overriding' are too ambiguous or inadequate. Strictly speaking, a small fine, for example, does not 'override' people's choice and they remain free, in a sense, to engage in any activity they would like to do. But it is still paternalistic. A paternalistic intervention would thus be better defined as 'taking steps to influence or alter people's choices for their own good'. And this happens, according to Sunstein, because the 'government does not believe that people's choices will promote their welfare' (Sunstein 2014, 54). A paternalistic intervention is meant to protect people against themselves—here, against the consequences of the choices that they are expected to make. This can be done by overruling choices, but also by influencing them in some way.

It can be argued that this definition does not cover important cases of paternalism, such as those discussed by Haybron and Alexandrova (2013), or that it fails to make clear what exactly the concern with paternalism is (Hausman 2018). According to Haybron and Alexandrova, by not deferring to the multiple ways in which individuals may value the options open to them, one can act paternalistically towards them, even if the subsequent intervention does not influence or alter their choices in any way[7]. In any case, every definition of paternalism, as Shiffrin (2000) has pointed out, must make clear what is the normative significance of paternalism. In the context of traditional welfare economics, it seems that the controversy behavioural

[6]In the following, I will refer to these properties (in addition to completeness, transitivity) when saying that preferences are rational. More will be said about rationality in a strategic setting in the following.

[7]The definition I will use do not include this concern about pluralism. It is enough for my purpose that this definition gives a sufficient condition for identifying acts of paternalism.

paternalism has stirred is partly imputable to the fact that it contradicts the value of consumer sovereignty, which traditional welfare economics upholds[8].

We may define consumer sovereignty (or more broadly, individual sovereignty) as the freedom to choose according to any of the preferences we might have at the moment of choosing[9]. Granting this freedom to individuals who are not rational results in choices that may be judged as mistaken in light of what appear to be their 'true preferences' (Sunstein and Thaler 2008), because what is chosen is not what they truly prefer. By nudging people or by restricting their options to make them better off in light of these 'true preferences', a paternalistic intervention would contradict consumer sovereignty defined in this sense. This may be a concern for economists endorsing traditional welfare economics because of its links with the liberal tradition. We may say that an intervention is paternalistic, in the sense that is relevant for our purpose, if it makes it difficult or impossible for people to choose according to the particular preferences that they have at some moment of choice.

A second point of criticism of the argument is that a definition of paternalism should exclude public interventions intended to solve collective action problems. Addressing these issues by way of nudges, taxes, or even coercion is generally not considered as paternalistic[10], even if it would make everyone better off because the failure of individuals to cooperate or coordinate with other people is harming these other people besides the individuals themselves. Insofar as traditional market failures highlighted in welfare economics are the result of such a lack of cooperation or coordination, a public intervention intended to correct them does not fall into the class of paternalistic acts. It must therefore be explicit that the paternalistic intervention is meant to promote the welfare of paternalistically protected individuals exclusively. In light of these two points, the anti-paternalist argument can be rephrased in this way:

[8]As emphasized by Sugden: 'welfare economists often say that, in using preference-satisfaction as a normative criterion for assessing public policies, they are treating each individual as the best judge of his own welfare. In this sense, neoclassical welfare economics can claim to uphold the non-paternalism of the liberal tradition' (Sugden 2018a, 6). Behavioural paternalism breaks with this tradition.

[9]This is simply the idea that each individual is the best judge of his welfare (See Blaug 1992, 125).

[10]see Sunstein (2008): 'to the extent that [laws] solve a collective action problem, they should not be seen as paternalistic at all'. See also Le Grand and New (2015).

1. (Definition) A paternalistic intervention takes steps to make it more difficult or impossible for people to choose what they prefer (at some moment of choice), with the intention of making them (and them only) better off.

2. Individuals always choose what they prefer (at any moment of choice), and they prefer what is best for them.

3. Making it more difficult or impossible for people to choose what they prefer (at some moment of choice) cannot make them (and them only) better off.

4. (Conclusion) A paternalistic intervention is bound to fail.

The third point of criticism can now be made clear: even if premise (2) is correct and individuals are perfectly rational, the conclusion (4) is not obtained without the addition of premise (3). The remainder of the chapter will describe a class of counterexamples to premise (3), showing that making it more difficult or preventing people from choosing what they prefer can actually make them better off, even if they always prefer what is good for them. Under these conditions, an intervention intended to make them better off may well succeed: this is what I call 'paternalism for rational agents'. Since the justification of this paternalism does not depend on the truth of premise (2), it represents a form of paternalism distinct from behavioural paternalism, which only arises in economics once doubt has been cast on premise (2). Since, as will be shown later, it is the context of strategic interactions that make this paternalistic intervention relevant, I will call it from now on 'strategic paternalism'.

3.2 Means paternalism and ends paternalism

Advocates of the new behavioural paternalism often frame their preferred version of it as an instance of 'means paternalism'[11]. As Sunstein puts it, in acting paternalistically 'government might well accept people's ends but

[11]For example, 'we have no interest in telling people what to do. We want to help them achieve their own goals (...) We just want to reduce what people would themselves call errors' (Thaler 2015, 325) This idea is further developed in Sunstein (2014), and Le Grand and New (2015). Conly (2013) also endorses means paternalism.

conclude that their choices will not promote those ends' (Sunstein 2014, 61). The point of behavioural paternalism would thus be to help people to better achieve their ends, as stated in their 'true preferences'. By contrast, even the use of nudges, warnings, and others 'soft' paternalistic interventions would not respect people's ends if it would make it more difficult to enjoy, for example, casual sex or the kind of activities that people intrinsically value, but which would be condemned by certain moral doctrines. We would thus have a case of 'ends paternalism', which appears much more damaging to individuals' autonomy.

As a consequence, the first virtue of this distinction is to define a form of paternalism (means paternalism) as 'minimal' as possible—as Haybron and Alexandrova (2013) would say—, by limiting the domain of potential paternalistic interventions. These interventions would be acceptable only when they concern means that individual would choose to reach their ends. If individuals do not choose the most appropriate means to reach their ends, the government may step in to intervene. However, an intervention that interferes with their ends would need much more justification than what behavioural paternalists can offer.

Indeed, advocates of the new behavioural paternalism often start from a Millian position[12], according to which governments have no business interfering with 'self-regarding' choices that do not harm others (or from the 'normative minimalism' of traditional welfare economics), and propose to depart from it by referring to the findings of behavioural economics. We are told that reasoning failures, biases, and problems of self-control prevent people from choosing the most appropriate means to reach their own ends ('as judged by themselves'). According to Sunstein (2014), the most that behavioural economics can do, normatively, is to back up this departure from the Millian position—but nothing more. That is why behavioural paternalism should limit itself to interfering with means people choose, but never with ends. In practice, behavioural paternalists[13] recognise the difficulty of drawing a sharp line between means and ends. Reasoning failures are key here, because insofar as the humean dictum that reason is the slave of passions is true[14], and thus reason is essentially instrumental, we are assured that intervening

[12]This is in particular what Sunstein (2014) does.

[13]see Sunstein (2014), Le Grand and New (2015).

[14]This is the point defended by Le Grand and New (2015): 'To intervene in this "reason" would be means-related paternalism; to question the "passions" themselves would be ends-related'.

to correct reasoning failures avoid the pitfalls of ends paternalism. Hence the insistence of behavioural paternalism on correcting 'mistakes', even if some other sort of behavioural 'anomalies', such as problems of self-control, are more difficult to describe as mistakes, as emphasized by Sunstein (2014).

The same framework of means paternalism can be used to describe the strategic paternalism that I will define and exemplify in the next sections, although strategic paternalism does not involve correcting any mistake. Since simple game theoretic models show that individuals can be made better off if we interfere with some of their own choices, a departure from the Millian position (or the 'minimalist' position) is similarly warranted. Game theory makes it easy to distinguish between means and ends. A game is described by each agent's set of strategies, a set of outcomes associated with each possible profile of strategies (a profile of strategies being the collection of the strategies chosen by each agent), and the preferences of the agents over these outcomes. Outcomes can be described either in terms of material rewards (such as the amount of money that someone wins at the end of the game) or as encompassing broader aspects of the final situations of the game. If we consider that individuals are rational when they only care about material incentives, we may describe the outcomes in terms of monetary payoffs and assume that the individuals always prefer to have higher payoffs.

In terms of means and ends, the preferences that an agent has over outcomes can be considered as expressing her ends, and the set of strategies includes all the means available to the agents to reach their ends. In a prisoner's dilemma, for example, the strategy to cooperate or to defect has no value beyond its ability to better satisfy the agent's preferences over outcomes. It is only a means to an end, the end being to get the outcome that the agent prefers the most among those available, given the choice of strategies of the others. One can thus say that an intervention has helped someone to reach her end—and has made her better off—if it has enabled her to reach an outcome which was previously unavailable for her, and which is better ranked in her preference ordering.

In a game theory framework, individuals do not choose directly an outcome, but a strategy. The outcome that they get depends on the choices of strategies of other agents. Therefore the notion of rationality must be completed to define how someone can have preferences over strategies that are derived rationally from their preferences over outcomes—assuming that these preferences are themselves rational. In a dynamic game with complete and perfect information, an important solution concept that I will use in

the following is that of a subgame perfect equilibrium. One intuitive way to justify that agents will choose the strategies that, taken together, constitute a subgame perfect equilibrium is that they use backward induction to determine which strategies the other will choose.

Consider the case of a game between the owner of the unique grocery store of a small village, confronted with the potential competition of another seller who may settle in the village. In the first period of the game, the latter (the entrant) chooses to settle or not in the village. In the second period of the game, the former (the incumbent) chooses to retaliate by lowering prices—and selling at a loss—or do nothing. As the entrant knows that, if she were to enter, the incumbent would be better off doing nothing, she can take for granted that the strategy to enter will give her the regular profit that she can expect to make in the village in a duopoly situation. As she prefers this outcome to the situation where she stays out of the village (but not to the situation where the incumbent fights back), she can infer that entering is the best strategy to achieve her goals. The resulting situation, in which the entrant chooses to enter and the incumbent chooses to do nothing, is the only subgame perfect equilibrium of the game.

This process of reasoning backwards is relatively straightforward but can be cognitively demanding for agents and run into multiple problems. For our purpose, however, we can note that in simple situations like these, which cover the cases that will be presented in the following, backward induction leads to the same result as the iterated elimination of (weakly) dominated strategies[15]. Since our purpose is to show that a rational agent may be the target of a genuinely paternalistic intervention, and since it is the incumbent who will be of concern in the following, we only need to pay attention to the rationality of the incumbent, at the time where he chooses his strategy.

In dynamic games, strategies can be thought of as plans of action, contingent on the choice of actions of others. For the incumbent, the contingent plan where he fights back if the entrant enters can never make him better off than the plan where he reacts to this entrance by doing nothing. Hence the former strategy is (weakly) dominated by the latter. All we need to assume is therefore that it would not be rational for him to choose such a weakly dominated strategy. In game theory, agents are rational if they choose the

[15]More generally, the two are equivalent provided that we choose the right order of elimination of weakly dominated strategies that eliminates all strategies but the subgame perfect equilibrium of the game.

strategy that constitutes the best response to their probabilistic beliefs about the strategy chosen by the others[16]. Since the only set of beliefs that would make the agent plan to fight back if the entrant enters puts zero probability on the event that the entrant chooses to enter, it can be ruled out on the assumption that such an event is known by the incumbent to be possible[17]. The preference of the incumbent for the strategy where he does nothing if the entrant enters thus derives from his preferences over outcomes and the belief that such an entrance is possible.

In sum, an individual is rational, in the context that will be discussed in the next section, if he has stable, context-independent preferences defined over outcomes (which can take the form of monetary payoffs) and chooses the strategy that constitutes the best response to the beliefs that he is justified to have about other agents' behaviours. But as we will see, that does not imply that the strategy which is rationally chosen is the best means to further his welfare (reach a higher payoff). Means paternalism may then come into play and turn things around.

3.3 Thomas Schelling and the logic of strategic commitment

How is it possible that people end up worse off when they have more options, given that they are rational and choosing what is best for them? As Dixit and Nalebuff explain: 'in single-person decisions, greater freedom of action can never hurt. But in games, it can hurt because its existence can influence other players' actions' Dixit and Nalebuff (2008, 54). This result, well-known in game theory, is sometimes called the paradox of strategic commitment. Since the outcome of a game is jointly determined by every player's choice of strategy, any change in the parameters of the game is likely to influence how

[16]The best response is the one that maximizes the expected utility of the agent, given the payoffs associated to the outcomes and his beliefs about other agents' behaviours.

[17]It is also necessary to suppose that the entrant will correctly anticipate that the incumbent is rational in this sense, and react by entering. But the exact mechanism by which this happens can be left undescribed since only the rationality of the incumbent matters for my purpose. It will be more convenient to assume, however, that both agents are rational and playing a subgame perfect equilibrium. In this case, we have to assume that rationality is common knowledge, which implies that each agent knows that the other is rational, and knows that the other knows that he himself is rational, and so on *ad infinitum*.

other players play, and therefore to change the nature of the outcomes that an individual can reach by choosing one of his strategies—making this player potentially better off. This results in a falsification of premise (3) in the anti-paternalist argument presented in the first section. Making it more difficult or impossible to choose a strategy may change other players' expectations about the action of the individual, which results in changing the outcome of the game, potentially making him better off.

Let us go back to the case discussed earlier where a seller (the incumbent) is confronted by a potential competitor (the entrant). Since the choice of the entrant whether to enter or not depends on her expectations about the incumbent's reaction to this choice, the incumbent may want to make an announcement before the start of the game, threatening the entrant to fight back if she enters. If the entrant believes this announcement, she will stay out and the incumbent will be better off. But every game theory textbook warns that such an announcement is not credible, since the entrant knows that if she were to enter, the incumbent would be better off doing nothing. In game theoretic terms, the situation where the entrant chooses to stay while the incumbent chooses to fight back if he enters is not a subgame perfect equilibrium.

A threat or a promise lacks credibility, according to Kreps, when '*ex ante* incentives to make the promises or threats do not match the *ex post* incentives to carry them out' (Kreps 1990, 52). This kind of situation is very common, and as noted by Kreps, was already well identified in the early days of game theory (notably by Stackelberg in the 1930s). Game theory's contribution to the analysis of credibility problems is that it makes clear in what way the rationality of the agent is necessary to generate such situations. The fact that the incumbent has no incentive, *ex post*, to carry out his threat is reflected in the fact that, *ex ante*, plans of action that involve fighting back if the entrant enters are (weakly) dominated, as we have seen. The announcement that he will choose a dominated strategy simply cannot be taken seriously if the incumbent is known to be rational. We can also note that the mismatch between incentives generates a mismatch between means and ends: because of his rationality, the incumbent cannot carry out the plan he has an incentive, ex ante, to formulate, and which would produce a better outcome.

Now suppose that the incumbent could, before the game starts, publicly commit to lower its price if the entrant enters. There are many ways (sometimes called 'commitment devices') to do so: he may give up the possibility

of doing nothing, or raise its cost to change his incentives—for example by signing a contract with a third party which would punish him if he does nothing[18]. That is what popular game theory textbooks recommend their reader to do to solve the credibility problem. Another possibility, less explored in game theory, involves renouncing rationality and, in particular, relying on the 'strategic roles of the emotions'. According to this line of reasoning, 'being known to experience certain emotions enables us to make commitments that would otherwise not be credible' (Frank 1988, 5). If the incumbent had a reputation for anger, it could persuade the entrant that he would ignore the incentives and fight back if the entrant entered. This may fit the game-theoretic framework described earlier if preferences over outcomes are understood to reflect only the material incentives of the situation. The power of anger is that it may make the incumbent disregard the material incentives he has to do nothing, and instead hurt himself in an outburst of rage. If this disposition is sufficiently known, it would make the incumbent's announcement credible.

What may be surprising is that choosing to commit oneself, in any of the ways I just described, seems bad for the agent, since commitment devices or strategies make it difficult to choose according to one's preferences at certain moments. Paradoxically, having options and being inclined to use them may be a disadvantage in situations of strategic interactions where people are rational. If someone is rich enough, his refusal to pay for things someone else wants him to buy is not credible. If someone is good at her job, her refusal to reach a certain productivity target if her job is on the line is not credible, etc. Having options and being rational makes individuals vulnerable to various forms of exploitation. People with bad intentions may even extort goods or services from them at no cost, with the power of a simple threat[19]. If they anticipate correctly some individual's behaviour, they may get what they want from him because, being rational, his actions will match their anticipation. The only way out is to get rid, publicly and irreversibly, of the power or the disposition to choose certain options. This move is called a strategic commitment.

However, this way of describing the paradox of strategic commitment,

[18]Elster (2000) mentions two other ways to commit by altering the material conditions of choice (rather than renouncing rationality): make options available with a delay or insulating oneself from knowledge about their existence.

[19]The logic of a threat, as underlined by Schelling (1960/1980), is that consequences only need to be carried out only if the threat is not successful (contrary to the promise).

which is standard in game theory textbooks[20] does not do justice to the first systematic description of the paradox, which comes from Schelling (1960/1980). Schelling did not imply that the gain in welfare obtained by the individual has necessarily to come from her own initiative—her decision to commit herself to a course of action. Indeed, it is not necessary for the logic of the paradox, as it was described earlier, that it shall be the individual herself who limits her own options. It can well be a benevolent third party who decides to remove someone's options for their own good: this is precisely the paternalistic intervention that this chapter vows to explore. Tellingly, Schelling never spoke of a 'paradox of strategic commitment' in his book but of a 'paradox of strategic advantage', an expression that does not mention the source of this advantage.

Consider this example from Schelling:

> 'An old English law that made it a serious crime to *pay* tribute to coastal pirates does not necessarily appear either cruel or anomalous in the light of a theory of strategy'

And later:

> '[Game theory] helps explain why a sufficiently severe and certain penalty on the *payment* of blackmail can protect the potential victim' (Schelling 1960/1980, 158)

To give a little more details, suppose that the coastline of a country is infested with pirates. If the government makes it a 'serious crime' to pay tribute to these pirates, it enables coastal travellers to *credibly* refuse to pay, since they would have to incur a very large penalty from the state if they were to comply with the pirates' threat. Pirates no longer have any reason to carry on their activities since their goal (being paid tributes) is now unreachable. If everyone is correctly anticipating everyone else's actions, piracy eventually disappears without anyone having to incur the penalty fixed by the government. Hence the idea that it is not at all 'cruel or anomalous', from a game-theoretic point of view, to make laws aimed at punishing potential victims.

What is remarkable in this example is that (1) the implementation of this law makes it impossible (or very difficult) for coastal travellers to choose

[20] see Fudenberg and Tirole (1991), Kreps (1990).

what they prefer when threatened by the pirates; (2) This law is meant to improve the situation of travellers (and only them) from their own perspective ('as judged by themselves'), and thus it respects their ends, since it is intended to eradicate or limit piracy, which is the best outcome of this strategic situation; (3) the intervention does not aim at correcting a failure in the decision-making process of individuals, as 'nudges' or bans are intended to do according to behavioural paternalism; coastal travellers are supposed to be rational in the sense defined earlier; (4) it is crucial that coastal travellers have no power to avoid the imposition of a fine and even that they cannot organize themselves politically to withdraw the law, because this would undermine the credibility of their refusal to pay tribute and encourage the pirate to continue their activities.

We thus have a situation where an intervention is making individuals better off, according to their own standards. And the fact that this intervention prevents people from doing something that would not harm anyone but themselves (paying tribute to the pirate) suggests that we have here a case of paternalism[21]. Points (1) and (2) stated above show that the definition of paternalism given in the first section is satisfied and that we have a case of means paternalism. By hypothesis, the intervention does not interfere with the ends of the individuals, which is to avoid paying tribute to the pirates.

Beyond this specific example, the logic of the argument developed in this section shows that the intervention of benevolent third parties can greatly improve the situation of individuals when they cannot credibly commit themselves to follow a course of action that would be beneficial for them. Since, as Schelling suggests, losing some options (or being less disposed to choose them) publicly and irreversibly is key to establishing someone's credibility, this intervention appears even more effective than the more familiar strategic commitment described earlier, because the source of the commitment is the action of a distant third party and not the individual themselves. I propose to call the class of such intervention 'strategic paternalism', to distinguish it from other cases of means paternalism.

[21]One might object that the intervention is not necessarily paternalistic because the payment of tributes incites the pirates to continue their activity and therefore it harms (indirectly) other people. But the payment of tributes does not necessarily generate externalities on other travellers because the activity of the pirates can be completely independent of past tributes (e.g. pirates do not save and invest, do not let newcomers become pirates, etc.)

3.4 Why strategic paternalism?

The intervention described before would count as paternalistic, according to the definition given in the first section, because it makes it impossible or difficult for people to choose what they may want at some point in time, to make them (and them only) better off. The form of paternalism I am describing here is about protecting individuals from themselves, which implies some sort of failure on the part of individuals. A rational individual who may want *ex ante*—for good reasons—to announce that he would not pay tribute to the pirate if attacked is exhibiting such a failure, as he cannot commit to making this announcement credible, and his actions would contradict it. This failure would simply not exist if the individual was (known to be) irrational and prone to fight back even when it is not advantageous to do so. Rationality, and the predictability it confers to the actions of individuals, can be seen as a disadvantage in certain contexts, as it prevents individuals from getting what they want[22]. The individual fails to take the appropriate means to reach his end: he fails to take actions to credibly commit and carry out the plan as announced. The only way he could commit himself if he did not have a commitment device would be to renounce rationality, which he cannot, by assumption, really do.

Why, then, would he not choose to use a commitment device? Why is the paternalistic intervention I described needed? I have argued in the previous section that individuals themselves do not need to limit their options to gain a strategic advantage. It could be done by a third party, to the same effect. This raises the question of whether a third party *should* limit the individual's option, rather than the individual herself. Rational individuals endowed with enough information would presumably recognize the need for a commitment device, and use it if available, which would make the intervention unnecessary. Irrational individuals, by contrast, may not spontaneously recognize their own failure to achieve their goal, which makes the case for a paternalistic intervention. There are many reasons why a third party, and in particular public authorities, may be in a better position to confer to individuals a strategic advantage and establish the credibility of their course of action: (1) because the commitment needs to be *public* to be effective, public authorities are more able to publicize the fact that someone has lost some

[22]If the interaction between the pirates and the travellers were repeated, fighting back or not paying the pirate could become rational, in the sense defined earlier. It is assumed here that such a repetition does not occur.

of her options and to make it common knowledge ; (2) because the commitment needs to be *irreversible*, public authorities are more able to refuse the demands of individuals to renounce their commitment and undo what they have done. As Schelling has pointed out, it is crucial to establish the credibility of someone that they have no say in the implementation of the procedure that limits their options. Suppose that coastal travellers could have a say in the implementation of the law punishing them if they pay tributes to the pirates or its application. It would create a bargaining opportunity for the pirates because it would mean that the decision to pay tribute is still in the hands of the travellers. Pirates might be able to pressure coastal travellers into repealing the law or preventing its adoption, or being exempted from its application. Only if lawmakers are independent of travellers can a strategic advantage be conferred on them.

However, this does not answer the following crucial question: even if we admit that a third party or public authorities are better placed to establish someone's credibility, why not provide individuals with a commitment device and leave it up to them the choice to use it? Provided that this choice is irreversible, that would do the trick: there would be no justification for a paternalistic intervention that restricts individuals' options without consulting them when they would be willing to do it themselves. I will now show why this is not generally true. Let us consider the case of a fundamental feature of the voting process in democracies, brought up by Schelling in his book as an example of the paradox of strategic advantage:

> What is the secret ballot but a device to rob the voter of his power to sell his vote? It is not alone the secrecy, but the *mandatory* secrecy, that robs him of his power. He not only may vote in secret, but he *must* if the system is to work. He must be denied any means of proving which way he voted. And what he is robbed of is not just an asset that he might sell; he is stripped of his power to be intimidated. He is made impotent to meet the demands of blackmail. There may be no limit to violence that he can be threatened with if he is truly free to bargain away his vote, since the threatened violence is not carried out anyway if it is frightening enough to persuade him. But when the voter is powerless to prove that he complied with the threat, both he and those who would threaten him know that any punishment would be unrelated to the way he actually voted. And the threat, being

useless, goes idle[23]. (Schelling 1960/1980, 148)

Laws requiring secret ballots are a case of strategic paternalism. Secrecy is not only meant to protect democracy from the consequences of buying and selling votes but also to protect individuals themselves from the violence and threats that they might be facing without a secret ballot if they vote the wrong way. Lawmakers have made secrecy mandatory when it could be optional: we could be provided by the electoral rules the *option* to use secret ballot or not. Citizens could tick a box in a form when they register to vote, stating that they do not want their vote to be publicly disclosed. This option would offer a commitment device to rational, well-informed individuals. Why is this option not generally provided by electoral rules? It may be that lawmakers think that individuals are not rational enough, or well-informed enough, to make the correct decision for themselves. But a more direct explanation is the mere fact that the power to threaten someone is not limited to the choice of a ballot. People can be threatened with reprisals if they do not vote the right way *and* refuse to make their vote public. The threat may also relate to the choice to use a commitment device. Whenever there is choice, there is a possibility of being threatened. The asymmetry of power is not the only reason why people are vulnerable to threats: as Schelling argues, It is also the simple fact that they have some choice, and may be under pressure as a result. In such contexts, strategic paternalism appears indispensable. Every person exposed to such threats is better off if he has no choice at all in matters that relate to the threats.

In essence, all this boils down to the following question: who has the power to make a strategic move? Schelling defines this concept as follows:

> A strategic move is one that influences the other person's choice in a manner favorable to one's self, by affecting the other person's expectations on how one's self will behave. One constrains the partner's choice in constraining one's own behavior. The object is to set up for one's self and communicate persuasively to the other player a mode of behavior (...) that leaves the other a simple maximization problem whose solution for him is the optimum for one's self, and to destroy the other's ability to do the same. (ibid., 160)

[23]Emphasis is not mine.

A strategic move, which gives someone a strategic advantage, is not a move that changes someone's position in the game, but a move that changes the game itself. The power to change the game to one's advantage is particularly costly to acquire, because, as the last part of Schelling's definition makes clear, it also implies preventing the other agent from doing the same—it can only exist as a privilege. Those who do not have this privilege, whose situation in the balance of power does not allow them to 'destroy' the other's ability to make a strategic move, are thus likely, however rational and informed they are, to suffer the strategic moves of others and solve a maximization problem that further these others' goals, not them. Powerful people who intimidate others and have a hold on them, on the other hand, can always destroy others' opportunities to make a strategic move by preventing them from using commitment devices. That would be the case of someone ordering others to vote in a certain way *and* to make their vote public so that he can punish them if they do not comply. Strategic paternalism solves this thorny problem simply by making secret ballots mandatory. Since in the contexts I have discussed, people are made vulnerable simply by having choices, a paternalistic intervention is warranted not despite, but because of its ability to remove choice. It intervenes in the balance of power by allowing people who are dominated by others, or people exposed to significant and extended threats, to gain a strategic advantage without having to make a strategic move (which they cannot do).

3.5 Beyond strategic paternalism

A paternalistic intervention, as I have defined it, takes steps to make it more difficult or impossible for people to choose what they prefer (at some moment of choice), with the intention of making them, and them only, better off. If this intervention helps people to take the most appropriate means to reach their ends—to make them better off 'as judged by themselves'—, it is a case of means paternalism. But there are two different ways to do it:

1. If the individual fails to improve her welfare directly through her own choices, an intervention intended to help her would be a case of behavioural paternalism.

2. If the individual fails to improve her welfare indirectly, through appropriate measures taken to influence other people's expectations about

her behaviour, an intervention intended to help her would be a case of strategic paternalism.

This paternalism is 'strategic' because it is relevant only when someone's welfare is affected by other people's expectations about her behaviour—which means we are in a situation of strategic interaction. In these situations, the government or another benevolent third party may try to alter these expectations by interfering with this person's own choice. I have described in the two previous sections a class of interventions that respond to this definition: those designed to establish the credibility of a course of action of some individuals by making some choices difficult or impossible. As credibility problems are widespread, this opens the way to a wide variety of interventions.

Strategic paternalism, as defined above, does not necessarily require that individuals should be rational. Bad things could also happen to an individual because other people expect her to act irrationally. If she is not considered sufficiently predictable or trustworthy, this may hinder her prospects, just as rationality and predictability, in the case of the paradox of strategic commitment, may make it harder to establish someone's credibility. However, the possibility that a rational individual might be the target of a genuinely paternalistic intervention is something that distinguishes strategic paternalism from its behavioural counterpart. It also contradicts the narrative according to which traditional welfare economics is immune from the sins of paternalism since game theory is part of its core and it makes strategic paternalism conceivable, as I have shown.

At this point, it is possible to suggest several directions in which debates on paternalism may be taken. The existence of two distinct brands of means paternalism calls for an integrated treatment of means paternalism, which would take into account both the direct and the indirect (strategic) effects of someone's behaviour on his own welfare. But by ignoring almost completely the second range of issues, advocates of behavioural paternalism have weakened their case, since the strategic effect of an intervention centred only on the direct effect of someone's behaviour on her welfare may be sub-optimal, as it ignores the broader strategic context. Suppose for example that by debiasing an individual prone to optimism bias, an intervention based on behavioural paternalism would make her less likely to overestimate her chance of success against a better opponent. The debiased individual, assessing her chance of success with lucidity, may choose (and be known to choose) to give in. This would make it harder for her to establish the credibility of some

possible contingent plan to fight back if the opponent steps in.

It is plausible that humans have evolved to become 'optimally irrational' (to use Lionel Page's book title[24]). A trait that generates irrationality in certain contexts, such as optimism bias, may somehow give an evolutionary advantage to its bearer and be useful in other (strategic) contexts. Evolution, in particular, enables people to make commitments simply by exhibiting their emotions, to show their underlying dispositions to choose or avoid certain things. As suggested by the example above, the various biases identified in behavioural economics are not unrelated to the capacity to make commitments and to bring them to the public attention. When behavioural paternalism proposes to debias individuals to make them more rational, strategic paternalism could, on the contrary, suggest pushing them towards a more brazen or even inconsistent approach to decision-making, or maybe not intervening at all because individuals are already, in the broader strategic context which is their own, 'optimally irrational'. An integrated account of means paternalism would thus try to determine if and when a paternalistic intervention is appropriate, in one sense or another. In particular, there cannot be a presumption that debiasing individuals is always a good thing. This integrated account has yet to be developed.

[24]See Page (2022).

Chapter 4

Normative Economics Without Preferences?

Opportunity Criterion, Anti-Paternalism and Preferences for Commitment

The results of behavioural economics have cast doubt on the idea that individuals' choices express stable, consistent and context-independent preferences. In a recent book, Robert Sugden (2018a) proposed to use an opportunity criterion to evaluate economic situations—according to which it is better to have more opportunity than less—instead of the traditional preference satisfaction criterion. Applying this criterion requires no information about individuals' preferences. However, the chapter shows how it conflicts with anti-paternalism (which Sugden endorses) when individuals have preferences for commitment—preferences for having less opportunity. The chapter also shows that a conflict between anti-paternalism and preferences for commitment exists even if we do not adopt the opportunity criterion. One thus cannot be a coherent anti-paternalist without attributing individuals a (minimal) preference for freedom[1].

Introduction

What criteria should be used to evaluate economic change? Different policies generate different outcomes, which can be judged as good or bad depending

[1] A version of this chapter is published in French in Badiei et al. (2022).

on the evaluative criterion employed. Traditionally, welfare economics begins by assuming that the individuals affected by these changes are capable of ranking each possible outcome according to their preferences. If one outcome is, in a sense, more preferred by individuals than another, the policy that generates that outcome is deemed better, from a welfare point of view. However, there is no guarantee that the various behavioural assumptions required to apply the criterion (including the stability and context-independence of individual preferences) are always satisfied. On the contrary, a multitude of results established by behavioural economics suggest that these assumptions fail to reflect the true behaviour of individuals—and this failure would be tantamount to undermining the foundations of welfare economics[2].

In the traditional approach to welfare economics, the economist is committed to recommending the policy that produces the outcomes that individuals most prefer. This principle is derived from what Haybron and Alexandrova call 'normative minimalism', which goes like this:

> It purports to keep value commitments to a minimum, if not to avoid them altogether, notably by orienting normative economics solely towards the satisfaction of preference, and thus (ostensibly) deferring to individuals' own value judgements. (Haybron and Alexandrova 2013, 159)

Because she strives to minimize her normative commitments, the 'minimalist' economist avoids imposing her own conception of what is good on individuals and therefore relies on them to know what is important to them. But this position, which can be described as 'anti-paternalist'[3], becomes empty if there is no way of identifying what individuals are supposed to prefer in a stable, consistent way—when, as observed in behavioural economics experiments, their actual choices cannot be rationalized by a stable, consistent preference relation.

One possible response to this problem is to try to reconstruct the preferences that the individual would have revealed through his choices, had he not

[2]Section 1.3 of Robert Sugden's book, The Community of Advantage (2018), entitled 'The Challenge from Behavioural Economics', presents and discusses four famous results from this literature, which lead to the questioning of traditional behavioural assumptions in economics. The two results that point to inconsistencies in agents' decisions are those that illustrate the importance of loss aversion and attention.

[3]Alexandrova and Haybron (2013) aim to show in their article, however, that the 'minimalist' economist actually fails to avoid paternalism.

been subject to the deviations from rationality that behavioral economists describe. However, it is not obvious that this approach, which Infante et al. (2016) describe as 'preference purification', can avoid making certain value judgments in order to identify those 'true preferences' (Thaler and Sunstein 2008) that the individual should have expressed under ideal conditions. Thus, this type of normative commitment runs the risk of being judged incompatible with economists' traditional anti-paternalism.

Robert Sugden's recent contribution to these debates, presented in the 2018 book *The Community of Advantage*, sets out to reconcile behavioural and welfare economics—or, as Sugden prefers to call it, 'normative economics'—and to adopt the latter's anti-paternalist commitment. However, this comes at the price of a serious revision, as Sugden's main proposal is to abandon the traditional criterion of preference satisfaction and launch a new 'normative economics without preferences' (Sugden 2021). The opportunity criterion he puts forward to replace the preference-satisfaction criterion can be described as 'the idea that is in each individual's interest to have more opportunity rather than less' (Sugden 2018a, 84). By broadening the set of opportunities available to an individual, we give him more latitude to make his choices and satisfy his preferences, whatever they may be. It is no longer necessary to identify the 'true preferences' of individuals to know that their situation has improved, from the point of view of this new criterion. It is enough to verify that they now have more opportunity than before. This would make it possible to dispense with any assumptions about the nature of people's preferences.

Moreover, adopting an opportunity criterion would mean avoiding making value judgements about what is good for individuals. By giving individuals more opportunity, we delegate to them the task of deciding what is good for them. The adoption of an opportunity criterion would thus be fully compatible with the anti-paternalist principle, while at the same time circumventing the problems raised by the existence of inconsistencies in the preferences revealed by individuals.

But what would individuals themselves think about having more opportunity? If we assume nothing about what individuals' preferences are, we cannot rule out the idea that they might actually prefer to have fewer opportunities than to have more. In particular, it is possible that an individual might prefer to keep his opportunity set as it is if adding an extra element to it would be a painful temptation for him to overcome, as Gul and Pesendorfer (2001) envisage. If we were to force this individual to make a choice

while being exposed to this temptation—for example, if we were to add a particularly tasty meat dish to a vegetarian's possible choices—we would be going against his preference. What would be the justification for doing so? The value judgment that would provide this justification could not appeal to this person's preferences, so it would impose the view that it is better for her to have more opportunity. This scenario suggests that it is impossible to maintain the opportunity criterion without making a value judgment about what is good for individuals—irrespective of what they actually prefer.

This chapter aims to describe precisely this general incompatibility between (1) the possibility that individuals sometimes prefer to reduce their opportunities—what Gul and Pesendorfer call a preference for commitment, (2) the anti-paternalist principle, and (3) the application of an opportunity criterion. This incompatibility is likely to raise a problem for a position like Sugden's since it entails either abandoning anti-paternalism by imposing on individuals a value judgment about what is good for them, or assuming the existence in individuals of a stable and consistent 'true preference' for having more opportunity (what I will call a 'preference for freedom'). The idea that it would be possible to do normative economics without preferences should therefore be reconsidered.

Sugden's exact position is, of course, complex and nuanced. On the one hand, the opportunity criterion he proposes is part of a broader 'contractarian perspective', in which the economist's recommendation is to set the terms on which citizens can agree to change their situation. The question of the role of paternalism and anti-paternalism in a contractarian approach remains, it seems, open[4]. On the other hand, Sugden's ambition in this book is essentially to defend the institution of the market, not the criterion of opportunity as a general ethical principle. That being said, Sugden is not alone in proposing an opportunity criterion. His work is part of a body of literature on freedom of choice that has developed over the last few decades[5], and which promotes this criterion independently of a contractarian perspective. Sugden himself has contributed to this literature, and one of his contributions (Sugden 2007) reveals the problem that will be discussed in greater detail here. The ambition of this chapter is thus to raise and discuss a problem that arises for any economist or philosopher holding the adoption of an opportunity criterion to be a response to the difficulties highlighted by behavioural

[4]See Rizzo and Dolde (2020) on this.

[5]A detailed review of this literature can be found in Barberà et al. (2004).

economics.

Section 1 sets out the definitions of the anti-paternalist principle, Sugden's opportunity criterion and preferences for commitment, which will be used in the following. In particular, in the context of normative economics, anti-paternalism is more than mere neutrality or abstention. It is a commitment to taking into account people's judgments about what matters to them. Section 2, which takes a closer look at how Sugden (2007) justifies his opportunity criterion, shows the incompatibility that exists between an anti-paternalist commitment, the application of an opportunity criterion and the possibility of preferences for commitment. The solution to an incompatibility of this kind lies in abandoning at least one of its terms; the implications of these various possible solutions are then discussed. In particular, it is not enough to abandon the use of the opportunity criterion to remove the incompatibility, for, as section 3 shows, the anti-paternalist principle can hardly be adopted consistently without attributing to individuals a (minimal) preference for freedom.

4.1 Definitional issues

4.1.1 The anti-paternalist principle

Welfare economics is based on a criterion of preference satisfaction, and the use of this criterion is consistent with a more general principle, which I propose to call anti-paternalism. According to Sugden, '[The criterion of preference satisfaction] was generally seen as embodying the principle that economists' recommendations should not be paternalistic: it was for each individual to judge what mattered to him or her' (2018, viii).

The anti-paternalist principle described here is not simply neutrality, in the sense of abstention from value judgments about what is good for individuals. It is true that the new welfare economics that took hold after the war, like its subsequent developments, owes much to Lionel Robbins' plea to banish value judgments from the field of economics, defined as a science[6]. But, as Sugden puts it, the anti-paternalist principle adopted by economists is inseparable from a certain value judgment, according to which the preferences of individuals (about what is good for them) have an intrinsic importance,

[6]See Baujard (2017) on this.

independent of their content. The principle embodies the idea that the individual is sovereign in the way he conceives his own good.

Why is this value judgment necessary? If the economist were to choose the recommendation he adopts at random—or even if he were to provide no recommendation at all—she would trivially respect the condition of simple neutrality, and in this she would make no value judgment. Because this is obviously not what is expected of the economist, the anti-paternalist principle must also include a condition that requires the economist to 'defer' to the value judgments of individuals (as mentioned in the passage quoted above from Alexandrova and Haybron), that is, not only to *not to impose* her own conception of what is good for individuals, but also to take account of theirs in her evaluation activity—to value what they value, all other things being equal. Because she is expected to formulate an evaluation of economic situations, the economist can only avoid imposing her own conception of what is good for individuals by basing her evaluation of situations exclusively on the judgments that individuals formulate about what is good for them.

A consequence of this idea is that people's preferences must have an existence independent of the possible procedures for eliciting them. If this were not the case, the principle would be irrelevant. One of the reasons why Sugden proposes replacing the preference satisfaction criterion with the opportunity criterion is that he doubts the independent existence of preferences. Indeed, '[Individuals] often come to decision problems without well-defined preferences that pre-exist the particular problem they face; instead, whatever preferences they need to deal with that problem are constructed on the course of thinking about it' (Sugden 2018a, 18).

But if these preferences were always a product of the decision-making process itself (and therefore context-dependent), there would be no hope of identifying what really matters to individuals, and the anti-paternalist principle would run on empty. There would be no way of ensuring that the use of this criterion of opportunity was in fact a commitment to anti-paternalism. It must be admitted that individuals are capable, at least in principle, of having certain stable and coherent preferences about what matters to them, for the anti-paternalist principle to have any relevance.

4.1.2 The possibility of preferences for commitment

Individuals do not just have preferences defined over the dishes in a restaurant, or over bundles of available goods, or, more generally, over possible

social states (called 'alternatives') that are the subject of economic evaluation; they may also have preferences defined over the 'menus' or 'opportunity sets' that contain all the alternatives from which they are required to choose. In particular, as Sunstein (2015) points out and illustrates, an individual may well prefer, in certain circumstances, not to choose. For example, a vegetarian may prefer being offered only a vegetarian dish rather than being offered a choice between meat and a vegetarian dish. A preference to avoid temptation is akin to a preference for commitment, defined in this sense. It is a preference based on opportunity sets, not directly on alternatives. There are many other reasons why a person would prefer not to choose, or at least to reduce the number of alternatives available to them:

- because making a decision is cognitively costly. Identifying the best alternative takes time and energy. The gain of being able to choose from a larger opportunity set can thus be more than offset by the cost of making the decision from a larger number of alternatives (Ortoleva 2013, Ergin and Sarver 2010).

- because the person who chooses one alternative thereby eliminates all the others, which can entail an emotional cost, particularly in terms of regret. Having more alternatives to choose from increases the likelihood of regret (Sarver 2008).

- because he prefers to delegate certain decisions to others, due to the fact that he wants to avoid the responsibility that would fall upon him otherwise (even if this decision would not be accompanied by regret).

- because, in a strategic context, it is sometimes preferable to limit one's options in order to gain a strategic advantage over one's opponent. When, for example, threatened by an opponent, it may be advantageous to give up one's options in order to give credibility to the refusal to give in to the threat (Schelling 1960/1980).

All these reasons justify or explain the existence of preferences defined over opportunity sets such that, sometimes, sets that provide fewer alternatives to choose from are preferred to those that provide more. A preference for commitment can thus be defined as the fact that, for at least one of the possible opportunity sets (let us call it S) from which an individual may have to choose, he prefers another set (let us call it T) offering fewer opportunities

than the first. The point I want to make here is not that this type of preference is particularly common, but simply that it exists. The fact that the evaluation criterion used (such as Sugden's opportunity criterion) does not require any information on individuals' preferences to be applied does not prevent the individuals concerned by the evaluation from having this type of preference. This is the source of the difficulty that will be presented in the second section.

4.1.3 The opportunity criterion

There are multiple ways to understand the idea that an opportunity set offers more opportunity than another. An entire literature[7] is devoted to defining and characterizing different rules for ranking sets in terms of freedom of choice. The type of opportunity criterion that is the subject of this chapter is more specific: it does not require any preference information to be applied. I will consider the formulation proposed by Sugden (2007) and taken up in Chapter 5 of his 2018 book, because it clearly raises the problem to be discussed—but it should not be forgotten that this problem is more general.

The question Sugden poses in this 2007 article is how to define an opportunity criterion when individuals' preferences are likely to change, especially over time. To answer this question, Sugden defines an extended opportunity criterion, which applies not only to opportunity sets, but also to choices among opportunity sets (and so on, indefinitely). For example, when a vegetarian chooses to order a meat dish or a vegetarian dish at a restaurant, he is making a choice from the menu or opportunity set provided to him by that restaurant. When, previously, this vegetarian had chosen his restaurant, it was as if he had chosen a certain opportunity set. And, just as it is possible to ask whether a certain opportunity set (associated, for example, with the choice of a restaurant) gives more opportunity than another, it is also possible to ask whether a certain *set* of opportunity sets (associated, for example, with the choice of a certain part of town where there are restaurants) gives more opportunity than another according to our extended criterion.

Sugden defines his criterion recursively. At the first level are the alternatives x and y. We can say that an alternative x weakly dominates another y, if, whatever may be the preferences of individuals, x is an alternative at least as good as y. The justification for such a judgment (which must be

[7] Namely, the freedom of choice literature, referred to in the introduction of this chapter.

independent of individuals' particular preferences) may be that any person, or any 'reasonable' person, would prefer x to y[8]. If x weakly dominates y, but the converse is not true, x strictly dominates y. If the converse is true, x and y are said to be equivalent. If neither x nor y weakly dominates the other, the two alternatives are said to be incomparable.

At the second level are the opportunity sets S, T. Here, Sugden defines a 'dominance extension' principle, according to which one set S weakly dominates another, T, if each alternative in set T is weakly dominated by at least one alternative in set S[9]. At the level of sets of sets, this dominance extension principle is itself extended in an analogous way: one set of sets U weakly dominates another, V, if each set of V is weakly dominated by at least one set of U. We thus have defined three dominance relations at three different levels. It is possible to define the principle in a similar way for higher levels of nesting, but this is not necessary for the purpose of this chapter.

This recursive definition enables us to evaluate, in terms of dominance, each different level: that of alternatives, that of opportunity sets and that of sets of opportunity sets. Sugden proposes to interpret this dominance relation in terms of opportunities. If an alternative x dominates another y, then x is a better opportunity than y. If one opportunity set S dominates another T, then S provides more opportunity than T, leading to the conclusion that S is better than T, according to the opportunity criterion. The reason why Sugden interprets the dominance relation in this way is rather intuitive: if S weakly dominates T, this means that for every alternative in T, there are alternatives that constitute opportunities at least as good in S. In other words, S is effectively richer in opportunities than T. And if all the alternatives are incomparable with each other, this opportunity criterion boils down to a simple inclusion criterion, according to which S provides at least as many opportunities as T if T is included in (or identical to) S. We thus have defined a criterion that produces a partial but fairly intuitive ranking, enabling us to compare opportunity sets independently of individuals' preferences for the alternatives.

[8]Sugden (2007, 666) interprets this dominance relation as the intersection of possible 'reasonable' preference relation.

[9]And S strictly dominates T if S weakly dominates T, but the converse is not true; if the converse is true, S and T are equivalent.

4.2 Conflicting principles

4.2.1 Incompatibility

The opportunity criterion defined above establishes comparisons between sets. Suppose, for example, that a set S offers more opportunity than another set T. The economist applying the criterion then produces an evaluation of these, and ranks S above T. Let us also suppose that an individual has a preference for commitment, as defined above: he prefers T to S precisely because T gives fewer opportunities than S, and thus enables him to commit himself. Now, if the economist recognizes the anti-paternalist principle, her evaluation must reflect that of the individual (assuming he is the only one concerned by the economist's evaluation). As a result, the economist must rank S above T—by virtue of the opportunity criterion—and T above S—by virtue of the anti-paternalist principle. The three definitions set out above are therefore incompatible: it is impossible to apply both the opportunity criterion and the anti-paternalist principle when dealing with an individual who expresses a preference for commitment.

One could respond that the point of adopting the opportunity criterion, as defined by Sugden, is that it avoids any reference to individual preferences. The opportunity criterion serves precisely to allow individuals to satisfy their preferences, whatever they may be, even if it is a preference for commitment in the sense given above, defined at the level of opportunity sets and no longer at the level of alternatives. An individual's preference for commitment could thus be satisfied if, at the level of sets of opportunity sets, he is given a choice between committing or not committing himself—a choice between S, which gives him more opportunity, and T, which gives him fewer. Leaving such a choice between S and T dominates, in Sugden's sense, forcing the individual to choose within the set S, and therefore gives him *more* opportunities. As a result, the fact that an individual has a preference for commitment would simply no longer be relevant. The program of 'normative economics without preferences' would thus be fulfilled.

And yet, the anti-paternalist principle is likely to conflict with the application of the opportunity criterion at all the levels that it is possible to define. Because at each level, individuals are likely to have a preference for commitment, which can be seen as expressing an evaluation contrary to that derived from the opportunity criterion, the anti-paternalist principle potentially conflicts with the application of the opportunity criterion at all these

levels[10]. The fundamental problem here is that the anti-paternalist principle invites us to ask, with regard to the evaluation criterion we choose, whether the preferences of individuals coincide effectively with the evaluations derived from the application of the criterion. Is the implicit value judgment produced by the application of the criterion imposed on individuals, or does it reflect their preferences? The answer to this question depends on the nature of these preferences. It is precisely because Sugden's opportunity criterion does not take into account agents' preferences (which may take the form of a preference for commitment) that it raises the difficulty formulated above.

The problem raised here appears in Sugden's own text. Let us go back to how he justifies his interpretation of the dominance relation in terms of opportunities. As we have seen, if S weakly dominates T, this means that, for each alternative in T, there are alternatives in S that are at least as good as in T. But at a higher level, that of the sets of opportunity sets U and V, how can we justify that 'U dominates V' on the basis of a judgment about the sets that U and V contain? We would have to put forward a justification analogous to the previous case: for every opportunity set in V, there are opportunity sets at least as 'good' in U. From Sugden's point of view, we only know that some sets provide *more* (or fewer) opportunity than others. How can we conclude that a set that gives more opportunity than another is also *better* than that other? This is a value judgement that is by no means universally shared. To justify the interpretation given to the principle of dominance extension, 'giving more opportunity' must also be 'better'—which could very well be denied. Sugden himself underlines this difficulty:

> We must assert that more opportunity is unambiguously preferable to less. That is a substantive normative claim, which not everyone will accept. I can only say that it is fundamental to my approach: Dominance extension expresses a commitment to the normative value of opportunity (Sugden 2007, 667).

In this passage, Sugden seems to be appealing to the readers, asking whether they too are ready to recognize 'the normative value of opportunity'. However, if we take anti-paternalism seriously, it is not the readers who are being asked this question, but the individuals concerned by the evaluation. Are they too prepared to endorse this value judgment about

[10]It is indeed possible to define a preference for commitment at each level.

what is good for them, as formulated by Sugden? This is not necessarily the case if they may have a preference for commitment. In this case, we cannot avoid imposing a value judgment on the individuals' good, which contradicts the anti-paternalist principle. This discrepancy in the justification of the opportunity criterion thus manifests the incompatibility described above.

4.2.2 Solutions

From the above, we deduce that it is therefore impossible to hold together (1) the anti-paternalist principle, (2) the possibility of preferences for commitment, and (3) the application of the opportunity criterion. As in every case of incompatibility in general, a solution arises when one of the terms producing the incompatibility is dropped.

Dropping the anti-paternalist principle

The first solution would be to accept the possibility of preferences for commitment, but still apply the opportunity criterion, which would mean dropping the anti-paternalist principle. The result is that the economist applies the opportunity criterion, even if it conflicts with the individual's own evaluation—who would judge that is it better to have less opportunity. A substantial value judgement is then imposed on individuals, which can be expressed in the terms of John Stuart Mill (1859/2006, 116): 'The principle of freedom cannot require that he should be free not to be free. It is not freedom to be allowed to alienate his freedom'. It is, of course, paradoxical to deny someone permission to renounce their freedom in the name of freedom itself. For Mill, it is a matter of justifying an exceptional paternalistic intervention, whereby an individual is prevented from significantly compromising his own future (Mill considers the case of an individual who would give or sell himself into slavery).

Sunstein (2015) uses the expression 'choice-requiring paternalism' to describe this type of intervention, which closes off certain present choices in order to keep others open in the future. There may be many reasons to resort to this type of paternalistic intervention in particular cases (notably in the extreme case of slavery contracts). In general, however, this solution seems necessarily unstable, because there is no *a priori* reason, from the point of view of a criterion of opportunity, to privilege future choices over present ones, rather than the other way around. Chapter 5 will explore this

issue in more detail and show how Mill's liberty principle can nevertheless lead to justifying such choice-requiring paternalism.

Dropping the possibility of preferences for commitment

The second solution consists in accepting the anti-paternalist principle and the application of the opportunity criterion, which implies ruling out the possibility of preferences for commitment. Unlike the previous solution, this one does not involve modifying normative principles to take account of reality, but rather adapting, in a sense, reality to principles. This solution thus consists in postulating a kind of pre-established harmony between the individuals concerned by the evaluation and the evaluating economist, since the criterion the latter selects would coincide exactly with the evaluations derived from the agents' preferences. One way of justifying this pre-established harmony would be to consider that agents have a stable and coherent 'true preference' for freedom and that the fact that an individual chooses to commit himself would only reveal that he has made a mistake. But apart from the fact that this justification breaks with the program of doing 'normative economics without preferences', it is hard to see how the attribution of such a 'true preference' for freedom could itself be justified. Yet this is the path Sugden does seem to take in the excerpt quoted above, which asks the readers to accept that more opportunity is better than less.

Another possible justification would be to restrict evaluations to situations where we can be reasonably certain that the individuals concerned have a preference for freedom. For example, as Sunstein (2015) points out, it is expected in a market context that individuals will actively choose among the available options—in particular, it is impossible to buy or sell anything without explicit consent, which means that the possibility of *not* buying or selling cannot be removed. We can therefore assume that individuals in a market would 'choose to choose', to use Sunstein's expression. But the attribution of this preference for freedom to individuals would then be based solely on the idea that they will act as expected in the institutional context in which they are. And there is nothing to suggest that, even in such contexts, some individuals would not prefer to limit their own choices.

Dropping the opportunity criterion

The last possible solution would be to accept the anti-paternalist principle and the possibility of a preference for commitment, but to abandon the opportunity criterion. The possibility that an individual has a preference for commitment would thus imply, in virtue of the anti-paternalist principle, that an opportunity set giving fewer opportunity than another is sometimes evaluated by the economist as superior to an opportunity giving more. Thus, the ranking of opportunity sets would in principle coincide with the preferences of individuals (defined over opportunity sets), insofar as these can be identified. If a vegetarian prefers not to be exposed to the temptation of choosing a meat dish, the anti-paternalist principle will judge the vegetarian restaurant as superior to the classic restaurant.

However, some of the other motivations mentioned above for a preference for commitment (cognitive cost of decision-making, regret, delegation) would lead to the avoidance of all choices, or of a whole range of choices. The person who wants to avoid regret at all costs, or to delegate his responsibilities completely, not only wishes to avoid exposure to a particular choice—which the anti-paternalist principle can take into account—but to choice in general. So what sense does it make to rely on such preferences? The problem is that, when an individual's preferences have this kind of structure, they come into conflict with the anti-paternalist principle itself. The result is an incompatibility between (1) the possibility of always preferring not to choose and (2) the anti-paternalist principle. The last section of this chapter is devoted to demonstrating this point.

4.3 A paradox of anti-paternalism

I will now show that accepting the anti-paternalist principle, as defined in the first section, means admitting that individuals cannot always prefer not to choose and that they therefore have a (minimal) preference for freedom. The argument to be developed in this section is that coherent anti-paternalism is only possible if individuals themselves are prepared to be treated anti-paternalistically, that is, if they value the freedom that anti-paternalism allows them to have. In short, anti-paternalism cannot be imposed on individuals without incoherence, and so a coherent application of the principle presupposes attributing a certain preference for freedom to individuals.

To make the argument as clear as possible, and to specify what kind of preference for freedom is implied by the argument, let us consider the simplest possible model of the relationship between an economist and the individuals affected by the policies she considers. The economist is supposed to produce an evaluation of the situation of individuals according to whether or not a certain policy is implemented, and then give a recommendation based on this evaluation. To keep things as simple as possible, let us assume that there is only one individual involved (which makes the model trivial, but nonetheless interesting from a logical point of view) and that only two alternative states x and y are to be evaluated (let us say that x is the state resulting from the implementation of the policy, and y the *status quo*). Let us also assume that there is no conflict of interest between the individual and the economist.

In addition, the economist is supposed to be able to elicit the individual's judgement about the two possible alternatives. Thus, the individual is supposed to indicate whether x is better than y or the opposite (assuming that the individual cannot be indifferent between the two). Let us say he indicates that x is better than y. This leaves the economist with two possible courses of action: either she accepts the individual's judgment, in accordance with the anti-paternalist principle, or she ignores this judgment and imposes her judgement instead. Finally, the economist's evaluation of the situation leads, following her recommendation, to the implementation of either x or y, depending on whether one or the other is deemed better by her (let us also assume that the economist cannot be indifferent between the two either).

The economist's recommendation therefore influences the individual's life. And, as a citizen, he also has an opinion about whether the economist should act anti-paternalistically towards him. Let us call a the situation where the economist adopts the individual's judgment, and b the situation where the economist ignores it. We can call a and b 'second-order' alternatives, to distinguish them from x and y, the 'first-order' alternatives. Since the individual judges x better than y if a occurs, x is recommended by the economist. If b occurs, the economist can recommend either x or y based on her own judgement of these alternatives. Since the individual is by hypothesis the only one affected by the first-order alternatives and second-order alternatives, we are justified in applying the anti-paternalist principle to these two pairs of alternatives.

Suppose the individual considers that b is better than a which means that he prefers that the economist *ignore* his judgment. In this case, applying the anti-paternalist principle to the alternatives x and y means that x is

considered better than y by the economist. But applying the principle to alternatives a and b means that y is considered better than x (if the economist so judges). The result is a contradiction: x is better than y and y is better than x.

One way of resolving the contradiction would be to argue that anti-paternalism does not apply to second-order alternatives. But, if the economist's attitude towards him is important to the individual, the anti-paternalist principle must also apply to these alternatives. The only way out is to reject the hypothesis made above. It must not be possible for b to be considered better than a by the individual. In other words, the individual cannot want the economist to ignore the way in which he himself judges the first-order alternatives x and y. Thus, endorsing the anti-paternalist principle requires the attribution of a certain structure to individual preferences. Such is the paradox of anti-paternalism. Just as you can only buy something from someone who wants to sell it, the anti-paternalist economist can only evaluate the situation of individuals in an anti-paternalist way if these individuals accept that their judgment is taken into account. The next stage of the argument is to show that individuals so disposed show a *de facto* preference for freedom. They value the freedom that the anti-paternalist gives them.

Indeed, since the individual is supposed to judge that a is better than b (i.e. he wants his judgment to be taken into account), this means that he values the situation where, when he decides that x is better than y, x occurs—following the economist's recommendation based on an anti-paternalistic evaluation. And when the individual decides that y is better than x then y occurs. If, on the other hand, the individual judged b better than a he would obtain x or y depending on the economist's judgment. Thus, when the individual prefers a to b he shows that he prefers the situation where he is free to choose between x and y (represented by the set $\{x, y\}$), to the situation where, independently of any action on his part, he obtains $\{x\}$ or $\{y\}$. It is therefore possible to conclude *a priori* that the individual prefers $\{x, y\}$ to $\{x\}$ or that it prefers $\{x, y\}$ to $\{x\}$, for x and y any possible alternatives.

The type of preference for freedom that has just been defined is *minimal*. Thus, an individual who seeks to avoid the temptation represented by the alternative x (the meat dish) would rank the opportunity set $\{y\}$ (the offer of the only vegetarian dish) above $\{x, y\}$ (the offer of both dishes), but also possibly above $\{x\}$ (the offer of the only meat dish), which does not contradict

this preference for freedom I have defined[11]. However, the case of an individual who refuses to rank $\{x, y\}$ above $\{x\}$ or $\{y\}$ contradicts this preference and renders the anti-paternalistic principle completely inapplicable. What has been shown is that it would make no sense to rely on the preferences of such an individual, who himself prefers to rely on others to decide what is good for him. Two attitudes are then possible. Either we judge this type of preference as pathological, which once again implies attributing a 'true preference', consistent and stable, for freedom to all individuals (in this minimal sense). But the program of 'normative economics without preferences' then appears out of reach. Or we abandon the anti-paternalist principle, which leaves open the question of the normative criterion to adopt, and of the type of justification to put forward to justify paternalistic intervention on the part of the economist.

Conclusion:

What should we think of the idea of a 'normative economics without preferences', based on the use of an opportunity criterion to evaluate economic situations? Offering more opportunity to an individual means ensuring that he or she has greater latitude to satisfy his preferences, without the need to identify what these individual preferences are. The use of this criterion thus holds out the promise of circumventing the difficulties raised by behavioural economics regarding the traditional approach to normative economics, while at the same time following the latter's anti-paternalism.

This chapter aimed to highlight some of the problems inherent in this program. Indeed, individuals may sometimes have a preference for commitment, i.e. a preference for limiting the range of their choices. The second section pointed out that, because of the anti-paternalist principle, such a preference conflicts with the application of the opportunity criterion. To remove this incompatibility, one possible solution is to dispense with the opportunity criterion. But the third section of this chapter has shown that, even in this case, the anti-paternalist principle cannot be applied consistently unless we postulate that individuals have a (minimal) 'true preference' for freedom. As a result, the possibility of 'normative economics without preferences' needs to be reconsidered.

[11] On the idea of a preference for freedom, see Puppe (1996).

Chapter 5

The Case Against Self-Constraint

Behavioural economics models and findings on self-control problems have provided the basis for the justification of paternalistic policies, which consider targeted individuals as incapable of solving these problems themselves. This new paternalistic program has triggered a significant backlash. In this article, I show how some of the arguments developed by anti-paternalist economists and philosophers also apply to the use, by the individuals themselves, of hard commitment devices (HCDs), which impose material penalties on individuals who fail to deliver on their commitment. HCDs have a disturbing character *as such*, that I propose to explain by connecting it to John Stuart Mill's famous argument against slavery contracts. This argument, once adapted, shows how a case can be made for the regulation of markets for HCDs from the perspective of freedom.

Introduction

Thomas Schelling was one of the first economists to think of problems of self-control as conflicts between 'impermanent selves, each in command part of the time, each with its own needs and desires during the time it is in command'. 'Self-management', or 'egonomics', to borrow Schelling's terms, is concerned with the art or science of 'coping with one's own behaviour as though it were another's'. As Schelling vividly describes, 'one of us, the nicotine addict, wants to smoke when he is in command; the other, concerned

about health and longevity, wants not to smoke ever, no matter who is in command, and therefore want now not to smoke then when he will want to' (Schelling 1984, 87). The subsequent behavioural economics literature on self-control that started with Thaler and Sheffrin (1981) has two implications, also underlined by Schelling. First, a person with a self-control problem who is aware of it (and thus described as 'sophisticated' in the literature) may be willing to pay for what is called a commitment device, that is, an arrangement that enables him to prevent his (anticipated) future self from taking a decision which he now considers inferior. Second, a person with a self-control problem who is not aware of it (and thus described as 'naive') may benefit from being prevented by a third party from taking a decision that the present self—and the third party involved—now considers inferior.

Forcing or influencing individuals to prevent them from making a bad decision at some moment is of course paternalist since it implies interfering with someone's choice for his own good. In particular, the existence of problems of self-control is a sufficient justification, according to Thaler and Sunstein (2008), to nudge the person to make the good decision identified by the so-called 'libertarian paternalist'. As O'Donoghue and Rabin (2003) claimed, 'economists will and should be ignored if we continue to insist that it is axiomatic that constantly trading stocks or accumulating consumer debt or becoming a heroin addict must be optimal for the people doing these things merely because they have chosen to do it' (O'Donoghue and Rabin 2003, 186). These positions have prompted a backlash from some anti-paternalist economists, who are often also at the same time defending the market as the best way to allocate goods, no matter how conflicted or flawed the individuals appear to be (Saint-Paul 2011, Whitman 2006, Sugden 2018a). These economists would oppose sin taxes or automatic (or forced) enrollment in saving programs, but they are often much less forthcoming on whether the use of commitment devices by the individuals themselves is good or bad—from the normative point of view they adopt. The chapter will argue that many of the arguments that they make against paternalistic interventions aiming at solving self-control problems can be rewritten as arguments against the existence of an (unregulated) market for commitment devices, which would go against their pro-market stance.

The goal of the chapter is to formulate a general argument against such markets for 'hard' commitment devices. A 'hard' commitment device (HCD) is, according to Bryan et al. (2010) definition, a commitment device 'that calls for real economic penalties for failure, or rewards for success' (the case

where an option is foreclosed can be interpreted as the situation where an infinite penalty would be attached to this option). By contrast, a 'soft' commitment device (SCD) is any device that has primarily psychological consequences. A classical example of an HCD is 'Christmas Clubs' accounts where individuals can deposit funds which are blocked until before Christmas, to prevent their impulsive self from overspending before Christmas (and thus not giving enough to their family). A Christmas Club account offers much less liquidity than a regular account, which makes it worse except for individuals with self-control problems. A classical example of SCD is the practice of mental accounting. For instance, someone would label a transparent box called 'money for Christmas', fill the box with money, and put it in a shared room for all to see. He would thus incur a psychological cost if he would withdraw from it.

I will depart from Bryan et al.'s definition in that I will only focus on costly commitment devices—penalties and not rewards are considered. As the Christmas Club example shows, with the use of HCDs people are either less free or worse off as a result of having committed themselves (and having paid for it), especially if they failed to deliver on their commitments. HCDs thus have a very disturbing character (which is not shared by SCDs). Take Schelling's example of a 'fat farm' where people agree to be forced to stay and exercise unless they reach a certain target in terms of weight: it would be perfectly justifiable to allow such an arrangement from the point of view of a social planner endorsing behavioural paternalism. At the extreme, even a slavery contract would be tolerable if it were designed to solve a self-control problem, and consented. The striking property of a HCD is that individuals can gain nothing by using and paying for them, apart from purported success in solving their self-control problems (which sometimes may be done in other ways, in particular by using SCDs). Besides, markets for HCDs enable firms to make a profit out of individuals with self-control problems, and not because, as is usually the case on a market, they sell better goods at a better price. If we combine this with the fact that markets are generally deemed responsible for generating issues of self-control among individuals[1], we get a sinister picture where private firms can at the same time supply the disease and the cure, each time cashing in on a profit at the expense of individuals' welfare and freedom.

[1]See for example historian David Courtwright's book, The Age of Addiction (2019), whose subtitle is: 'How Bad Habits Became Big Business'.

The goal of this chapter is thus to show from which perspective we can conclude that HCDs are bad *as such*. The question is not trivial because it is plausible that HCDs actually make some people better off, and their use is consented to by individuals, as I will explain in section 1. Two perspectives can be adopted. A certain representation of the agency of individuals needs to be adopted to make a judgement about the merits or demerits of HCDs. Approaches in terms of welfare usually rely on a certain version of a multiple selves model, which makes it hard to pinpoint why exactly HCDs are bad *as such*, as I will show in section 2. Another perspective is that of Sugden, who vehemently criticizes this model, suggests a different representation of the individual as 'responsible' and adopts an opportunity criterion to make normative judgements—which appeals to the value of freedom. But Sugden's opportunity criterion fails to assign a negative value to HCDs, in contradiction with his representation of the individual as 'responsible', as I will show in section 3. The fourth section will go further than Sugden and show how John Stuart Mill's argument against slavery contracts can be generalized, as philosopher David Archard reformulated it, to show that HCDs are outside the scope of Mill's liberty principle, without making substantial assumptions about psychology of individuals. The conclusive section suggests that SCDs, under the form of what Reijula and Hertwig (2022) call 'self-nudging', can provide a valuable alternative to using a HCD.

5.1 Antipaternalism and the markets for HCDs

Markets for HCDs can only exist because there is a demand for them. The agents of standard economic models, endowed with preferences which are stable and consistent over time, are never willing to pay for a HCD, because it is never useful to them. In particular, since inferior, suboptimal options are never chosen by them, they are indifferent between the situation where these options are present and the situation where they are removed[2]. As Bryan et al. (2010) point out in their survey, there exist three main kinds of behavioural economic models which imply that the agent represented in the

[2]The indirect utility criterion, which expresses the attitude of these agents towards opportunity sets, states that the value of a set if exactly the value of its best elements. As a result, removing suboptimal options from the set makes no difference to them.

model would be willing to pay for a HCD:

- **Hyperbolic discounting.** This kind of model, first outlined by Strotz (1956), shows how 'different selves differ in their assessment of the best course of action and consequently that each time's decision maker would like to restrict the set of choices available to his or her future selves' (Bryan et al., 676). An individual who discounts future flows of utility hyperbolically is necessarily time-inconsistent: he may prefer to receive ten euros in one month and one day rather than receiving five euros in one month, but take the five euros when the day comes to choose between taking five euros now or ten euros tomorrow. If this individual is sufficiently 'sophisticated' to anticipate that his preference will change in this way in the future, he may want to thwart the actions of his future self and make sure he will receive the ten euros, which makes him better off now.

- **Preferences for commitment.** Gul and Pesendorfer's (2001) model considers preferences over opportunity sets or 'menus'. If there exists a cost associated with being exposed to a tempting option, an individual may prefer to choose in a smaller set—possibly a singleton—than in a bigger one, which is formally equivalent to being willing to pay for a HCD that would remove certain options from her menu. A vegetarian would not want to be offered a meat dish in addition to her favourite vegetarian dish, because of the temptation it induces (even if she would choose the vegetarian dish). In this model, the agent consistently maximizes her total utility (which includes a 'temptation' disutility) when choosing menus, but her preference for commitment is due to some temptation which would be impossible to explain without the reference to an intra-personal conflict.

- **Dual-self models.** In these models inspired by Thaler and Shefrin (1981), a long-run, 'planner' self, concerned with the lifetime utility of the individual, has preferences which differ from one or multiple short-run, 'doer' self, making consumption decisions, and only cares about the present (he is 'myopic'). In Thaler and Shefrin's model, the planner is acting strategically and can either manipulate the doer's preferences to induce him to make the decisions that maximize the lifetime utility of the individual or alter the budget constraint of the 'doer' to produce the same effect. The latter is a case of commitment where the

doer is no longer free to consume as he wishes. One interesting aspect of this model is that it incorporates insights from psychology and in particular a differentiation between two different 'systems' of thought (Kahneman's systems 1 and 2)[3].

Indeed, these models can be understood as representing the economic agent as divided between two selves, who have 'two sets of preferences that are in conflict at a single point in time' (Thaler and Shifrin 1981, 394), even if usually only one self can make a decision at any point in time. A two-selves model is thus fundamentally different from a simple phenomenon of changing tastes and raises much more complex questions about the welfare of this individual. As we will see in the following, and as recognized by Thaler and Shrifin, thinking about multiple selves involves using 'organizational analogies', which give more explanatory power to the model, at the risk of losing sight of the unitary nature of the individual—something that cannot be lost without abandoning the idea of legal and moral responsibility.

For 'naive' agents unable to anticipate that their initial or optimal plan will be thwarted by their subsequent selves, self-control problems may result in overconsumption of food, alcohol, cigarettes, or undersaving, compared to their initial plan, or the plan that they might have made 'if they had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control' (Thaler and Sunstein 2008, 5-6). This gives an argument for a paternalistic public intervention that has the same effect as the voluntary use of a HCD. O'Donoughe and Rabin (2003, 2006) have explored the idea of 'optimal sin taxes', or 'optimal paternalism'. By overconsuming potato chips, the present self acts as if he is imposing a negative externality on the health of his future self. This analogy is reflected in the adoption of the term 'internality', adopted by many behavioural economists[4]. A 'sin tax' designed on the model of a Pigouvian tax would thus naturally lead agents with self-control problems to reduce their consumption to an optimal level, while not affecting other agents' welfare. The paternalistic characterization of these 'sin taxes' becomes somewhat blurred if one really takes seriously the multiple-selves framework—as Pigouvian taxes are not paternalistic at all. But from the point of view of the unitary agent, it falls into the definition

[3]See Thaler and Sunstein 2008.

[4]See Herrnstein et al. (1993). Sunstein (2015) even uses the term 'behavioural market failure'.

of paternalism given in chapter 3, as some choices that the individual would have done are made difficult or impossible[5].

Three types of arguments have been developed by economists opposed to such paternalistic public intervention:

- According to Cowen (1991), who develops here a theme from Schelling, when it comes to welfare evaluation, the literature on self-control (especially the literature on dual selves) tends to adopt uncritically the point of view of the long-run or planner self as representing the true interests of the individual. The short-run, or doer self is described as 'myopic' or 'impulsive', neglecting the fact that, for Cowen, he is the bearer of values of spontaneity, self-discovery, etc. Maybe economists and psychologists only adopt the point of view of the planner self because he is the only one deemed capable of acting strategically and considering the future[6]. But the doer self may also act strategically, according to Cowen. For example, some people would rush to answer calls for charity donations, because they know that their planner self, who is focused on rules and long-term goals, would not indulge in it when he is back in control, so to speak. Cowen pleads for a more balanced and complex vision of self-control problems, which contrasts the typical 'self-command' action of the planner with the necessity, in terms of self-management, of 'self-liberation'—the need to relax the sometimes excessive discipline of the planner self. In that perspective, paternalistic interventions almost infallibly favour the planner self, preventing self-liberation.

- According to Whitman (2006), the proposed paternalistic interventions are based on a vision of the interaction between selves in terms of internalities. But the inefficiency that results from the fact that some self does not internalize the consequence of his actions on others selves is not necessarily, or not adequately, addressed by Pigouvian 'sin taxes'. This paternalistic answer ignores the contributions of the Coasian approach to internality problems, which would not require a paternalistic intervention. A Coasian negotiation between selves is likely to be much more effective than outside intervention in addressing inefficiencies, even if

[5]See also Saint-Paul (2011) for a definition of paternalism that acknowledge this fact.

[6]Elster (1984; 2000) uses this as a criterion for identifying what he calls the 'authentic' self.

it does not necessarily result in the same kind of behaviour that the individual would have in the absence of self-control problems. The obvious fact that successive selves cannot really communicate between themselves does not prevent them from cooperating and making compromises, for example by following a clear-cut rule, as Ainslie (1992) described it. If this kind of cooperation takes place, an outside intervention risks perturbing the inner balance achieved by the cooperation of the selves and negatively affects the individual's welfare.

- Saint-Paul (2011) recognizes that paternalistic interventions may be effective in solving self-control problems in the short-term, but is worried about the long-term effects that the rise of the new behaviorally-informed, paternalistic style of government might have. Systematic paternalistic interventions, implemented each time a self-control problem is pointed out, involve a 'responsibility transfer' from the individuals to the state, the firms, or anyone who is considered to be 'unitary' enough—that is, not subject to self-control problems—to be able to cope with the consequences of other people's self-control problems. If unitary agents are required by the state to assist non-unitary agents or to accept to see their welfare restricted to do so, responsibility has a cost. This implies that the new paternalistic state is not incentive-compatible, since agents would want to avoid bearing the responsibility to assist others.

In the context of this debate, the role of markets is ambivalent: they give opportunities to the 'impulsive' self to overconsume, undersave, or simply evade the commitments already made by the planner self. Someone who has put funds in a Christmas Club account can simply go to the bank and get a credit to spend as he likes, undoing the plan of the planner self. This could justify either paternalistic regulations of markets or the creation of new markets for HCDs which would be impossible to evade—thus performing exactly the function that paternalists assign to their intervention. All that would be needed is to inform 'naive' decision-makers—unaware of the extent of their self-control problems—of the availability of these market solutions. Indeed, markets may offer HCDs in two different ways:

- Private producers can supply HCD unwittingly, as when a bank offers its client to buy assets that happen to be less liquid than others, which

ties up funds for some time and can be used by people with self-control problems to overcome their tendencies to 'overspend' at certain periods.

- Private producers can supply HCD as such, by designing products that enable individuals to make a hard (or soft) commitment. Thaler and Sunstein (2008) give the example of 'Clocky', a robot that wakes you up with an alarm and then runs away to force you to get out of bed to catch it and turn off the alarm. The doer self thus manages to get up on time, just as the planner self wanted. Less anecdotally, there now exists countless applications or programs that enable customers to commit themselves to reach a measurable target and pay a significant amount of money to the company selling it if they fail.

From the point of view of public intervention, the first kind of HCD is not as concerning as the second may be, since commitments made by using products designed for reasons other than dealing with self-control problems may be more easily evaded than commitments associated with the second kind of HCD, which are meant to be difficult or impossible to evade. What should we think of this market, in light of the debate outlined above? For antipaternalists, HCDs may represent an interesting compromise, as they enable individuals to solve themselves their commitment problems, without any imposition of sin taxes or responsibility transfers. The government may inform and even incentivize people to use HCDs instead of implementing paternalistic interventions. The usual arguments in favour of a decentralized market would apply here, as the selling of HCDs as private goods does not seem to involve any market failure.

The purpose of this chapter is to show that HCDs are bad as such, according to a freedom criterion, but not according to the traditional welfarist criterion. From the perspective of the 'liberty principle' that will be developed in section 4, markets for HCDs should be regulated so that the state only sanctions soft commitment device contracts and not hard ones. This conclusion only applies to commitment devices that aim at solving self-control problems. There are many reasons why people would sometimes 'choose not to choose'[7] and may want to commit themselves to follow some course of action. I will suppose that it is possible to discriminate between those reasons and that HCDs aiming at solving self-control problems are identifiable as such.

[7]see chapter 4 for a review of these reasons.

5.2 Three welfarist arguments against HCDs

The three arguments against paternalistic interventions evoked in the last section are all based on a welfarist perspective. They conclude that a paternalistic intervention would result, against its intentions, in making individuals worse off. But as we will see, these arguments can also be reformulated to be directed against markets for HCDs, which means that they are not anti-paternalists *per se*. They all depend on the multiple selves model in their formulation. The problems they raise lie in the fact that, when committing themselves, individuals deprive their future selves of their freedom, which prevents these selves from making the best of their situation. This absence of flexibility may thus be detrimental to the welfare of the individual as a whole. In this section, I will present three welfarist arguments against HCDs inspired by the anti-paternalist stances of Cowen, Saint-Paul and Whitman, and then explain why they cannot conclude that HCDs are bad as such, which suggests another perspective is needed.

5.2.1 The symmetry argument

Economists and psychologists may be tempted to take the side of the long-run, planner self—assimilating the preferences of the planner self to the 'true' preferences of the individual herself—because they make the implicit assumption that there is a fundamental asymmetry between selves. The purpose of Cowen's paper is to show (mainly by examples) that this assumption is not warranted in general—because the short-run self may also behave strategically towards the long-run self, and because his preferences also matter to the welfare of the individual, even when they are not aligned with those of the long-run self. Recognizing this absence of asymmetry leads to see the art of self-management as implying 'the unleashing of forces in such a way as to create a complex but coordinated processus of personality growth' (Cowen 1991, 373), which seems to mean that public authorities would do better to focus on this broader 'personality growth' rather than on the limited interests of the planner self.

Creating a more balanced self-management would mean abandoning the 'command and control' approach which embraces the point of view of the long-run self. The underlying analogy implicit in the reasoning of economists and psychologists who put so much emphasis on the perspective of the long-run self is that of centralized planning, which leaves no initiatives or flexibility

for agents in charge of executing the plan. Flexibility is not needed if one believes in the fundamental asymmetry between the selves. But if, on the contrary, the short-run self is the bearer of values of spontaneity and self-discovery, and his interests are as respectable as those of the long-run self, HCDs may have a negative value from the point of view of individual welfare since, being implemented by the long-run self, they fail to leave enough flexibility to the short-run self. HCDs would make it impossible to use certain techniques of self-liberation that short-run selves have at their disposal when they are not constrained by the planner self.

For example, from the point of view of the long-run self, the possibility of making sports bets or buying lottery tickets may be undesirable, from his own assessment of risks and benefits (the long-run self knows that the expected benefit is lower than the price of the lottery ticket). But using a HCD to prevent the short-run self from buying them may be a bad self-management practice, as the *possibility* to participate in the lottery or to make bets gives the individual the hope (the dream?) of improving his lot and thus make his present situation tolerable. Using a HCD, just as being the target of a paternalistic intervention prohibiting bets or lotteries, would jeopardize the coordination between selves which Cowen sees as necessary to reach 'personality growth'. Since the argument against paternalistic intervention is based on the benefit of leaving flexibility to the short-run self, it can be rephrased as an argument against the market for HCDs, which would give an undesirable advantage to the long-run self, who is too focused on discipline.

5.2.2 The information argument

Someone using a HCD anticipates a change in his preferences which would lead her, if she acted upon them, to get a result which is inferior according to her present preferences. But the fact that a decision makes someone better off or not has nothing to do with the moment where it is evaluated, and everything to do with the information available when it is taken. If information is not perfect, and preferences are not mere tastes but depend on the information available when they are formed, it becomes crucial to know which self is the most informed about the welfare implications of the actions of the doer self. For the impartial observer trying to evaluate the welfare of the individual, the question becomes: does the self willing to use a HCD really know what he is doing?

The planner/doer dichotomy is once again driven by a misleading analogy,

according to Daniel Read (2006). It is justified to be on the side of the self who wants to use a HCD only if he is capable of anticipating correctly the preferences of the ulterior selves and making the relevant trade-off, just as an ideal planner would do. But this is not the right way to understand the 'egonomics' of the individual, because preferences are formed according to contextual information, which only the self situated in the right context can apprehend. The person who commits herself to running each morning before going to work probably underestimates the pain that her ulterior selves will endure each morning, for a very long time. As Bryan et al. (2010) point out, Kahneman et al. (1997) suggest that 'pain is remembered differently from how it is experienced' (Bryan et al. 2010, 694), which would support the intuitive idea that 'pain becomes less memorable as time goes by', and therefore that the planner self is not in a good position to correctly evaluate the disutility of a future pain. Letting the prospective runner commit herself by promising to pay a significant amount in case she fails to exercise would be disastrous, as she would either lose money or deliver on her commitment at too high a cost.

This general argument against HCDs is, as Read (2006) remarks, similar to Hayek's knowledge problem, which was raised as an objection to central planning. Just as the central planner cannot get the right information—which is necessarily contextual and held by agents who have no means or incentive to communicate it (in the absence of a price system)—to make efficient allocation decisions, the planner self cannot gather now the information that will only be available in the future. This argument would lead to giving a negative value to HCDs, in the absence of perfect information, from the perspective of a welfare criterion. However, the task assigned to the planner self seems much less complex than that of Hayek's central planner. Even if it were not possible to make the precise intertemporal trade-offs that would justify committing oneself, the planner self could base his decision to commit or not on his (probabilistic) beliefs about the selves that will appear later. That decision would be justified from an expected utility view of welfare, even if it turned out to be wrong *ex post*. The previous argument may thus only make sense in a situation of radical or Knightian uncertainty, where it is impossible to define a probability distribution over possible selves. Why tie one's hand when the future is completely unknown? But individuals are not always, and maybe not often, in such a situation of radical uncertainty.

5.2.3 The incentives argument

In the Coasian approach of Whitman, the individual can overcome the internalities she is faced with, thanks to some form of intra-personal Coasian 'negotiation'. Since the parties involved cannot really negotiate, the cooperation between selves has to be some mutual acknowledgement, among selves, of each other's presence and importance. One example of such cooperation could be Cowen's example, mentioned earlier, of a long-run self accepting that the short-run self uses some amount of money to buy lottery tickets because it brings a hopeful perspective to the individual. Ainslie (1992) suggests that the implementation of such an 'agreement' among selves—which would seem impossible if no self can really ensure that the other respects his part of the agreement—take the form of a 'package deal'. A personal rule can be adopted, which is such that if a self, at one point in time, deviates from the rule, this deviation will be generalized, and the individual would thus end up in a situation so bad that every self would want to avoid it. One way to achieve that is to define 'bright lines' such that any small deviation from the rule would be acknowledged as a violation and rejection of the rule. For example, people would adopt a rule never to drink alcohol again, rather than a more flexible and convenient arrangement, because it is much more clear-cut and leaves no room for *ex post* rationalization and accommodation: the rule is either respected or violated, in which case the individual has lost his bearings and finds himself in a dangerous position. The agreement between selves holds, under these conditions, because all selves have a common interest in making sure that the rule is followed.

Insofar as the selves follow this kind of rules, the behaviour of the individual is 'unitary' and his choices are consistent and stable. What could prevent this agreement from being made? The Coasian approach suggests that this would happen when transaction costs are too high—the mechanisms by which such an agreement can be reached among selves are fragile, because, in the absence of a HCD, no external third party can enforce the agreement. But using a HCD would bring us back to a 'command and control' solution because HCDs are always used by one self to bind the others, which is totally at odds with the spirit of the Coasian approach. What is more, the possibility of doing with a HCD gives an incentive to the self in position to use it to give up on the process of a Coasian 'negotiation' and on reaching the subtle agreement to which it can lead.

What is crucial for the present argument is that if the process of inter-

nalization can be achieved by outside agents (private firms, governmental agencies), because HCDs are enforced, individuals are encouraged to delegate the task of managing their self-control problems to others, instead of doing it by themselves, through a Coasian 'negotiation'. The whole point of the coasian approach applied to a multiple selves framework is to explain how individuals can act consistently even if they have self-control problems. But if it is institutionally possible to delegate this task to others, there is no reason for individuals to invest in their own psychological capacity to overcome their intrapersonal conflicts and put it to good use. Such an evolution can paradoxically—and somehow, performatively—confirm the claim made by some behavioural economists that the 'paternalistically protected category of idiots' needs to be extended to include 'most people' (Camerer et al. 2003, 1218). If people do not have the incentives to avoid behaving like idiots, there is every reason to believe that they will. Moreover, if collective resources are used to assist people in overcoming psychological problems that they could—and could better—solve by themselves, instead of using them to build collective prosperity, some significant social loss will be incurred.

5.2.4 Intrapersonal prisoner's dilemma

Whatever may be the value of these arguments, they are not fit for my purpose, which is to account for the intuition that using HCDs is bad in itself. These arguments cannot conclude that HCDs are *always* bad because there exist at least one theoretical class of situations to which the three arguments cannot apply: intrapersonal prisoner's dilemmas (PDs). According to Andreou's description:

> Agents who discount future utility are fragmented into (...) times-slice selves. Each time-slice self is not indifferent to the fate of the other time-slice selves, but closer time-slice selves are favoured over more distant time-slice selves. Intrapersonal PDs exist when each time-slice self favors the achievement of a long-term goal but also prefers that the restraint needed to achieve the long-term goal be exercised not by her current self but by her future selves. (Andreou 2022, 6-7)

Undersaving problems have exactly these features: each 'time-slice' self would need to save a small amount to make sure the individual will get

a good retirement pension—which can be seen as a public good valued by each self. But at the same time, each self has an incentive to free-ride on the contribution of future selves and will do so if not forced to contribute. The resulting situation where no self contribute[8] is inefficient since every self prefers the situation where enough saving has been done. Under these conditions (1) every self would be willing to pay for a HCD forcing all selves to save the optimal amount of money, (2) there is no uncertainty about the payoffs faced by the different selves, as every self is in a symmetrical position and deeply care about the individual's welfare when retiring, (3) without a HCD, every self has an incentive to free-ride and the result would inevitably be undersaving. The possibility of intrapersonal PD shows that using a HCD is not necessarily a zero-sum operation: restricting one self's possibilities is not necessarily always only to another self's advantage.

Because of (1), the symmetry argument cannot apply to this case: every self is comparable to the others and shares the same interests. It would be bad, from every self's point of view, to be left with some flexibility. Because of (2), the information argument cannot apply either: the connection between the selves' savings and the retiree's welfare is straightforward and is not as distorted as the memory of pain and 'experienced utility' can be. Because of (3), the incentives argument cannot apply as a Coasian agreement between selves seems impossible to reach. The structure of a PD makes it necessary to punish non-cooperative behaviour to ensure that the optimum is reached. This cannot happen in such intra-personal conflicts unless a HCD is used. It would thus seem that intrapersonal PD provides the best possible case to justify the existence of markets for HCDs and paternalistic interventions if we adopt a welfare criterion. There is no way to preserve the retiree's standard of living other than to force each 'time-slice' self to save a sufficient amount of money while they can. It seems difficult to avoid the conclusion that something could be done to avoid undersaving problems, and the nature of the problem implies that every self would agree to be forced to save.

5.3 Sugden's responsible individuals

The three previous arguments fail to show that HCDs are always bad if our goal is to maximize individual welfare. The case of intra-personal PD pro-

[8]Or contribute only as much as its 'stand-alone' contribution, as in a classical public good game.

vides a compelling justification for the use of HCDs if we accept the multiple selves model which underlies this justification. As every self is made better off by using a HCD, no difficult normative assumption needs to be made about the weights that should be assigned to each self's set of preferences. We do not need to answer the difficult question raised by Schelling and Read, 'Which side are you on?'[9], because we can afford to be on every self's side. If individuals' intrapersonal conflicts take the form of a PD, protecting them from themselves would be warranted. However, if we follow Sugden's influential criticism of behavioural welfare economics—the literature aiming at reconciling standard welfare economics with the findings of behavioural economics—and libertarian paternalism, this conclusion is only the product of a representation of the economic agent which is particularly misleading.

According to Infante, Lecouteux and Sugden (2016), the fundamental flaw of these new approaches is to interpret the 'anomalies' pointed out by behavioural economists—such as contradicting one's earlier plans—as *mistakes* that individuals 'would not have made if they had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control' (Thaler and Sunstein 2008, 5-6). This interpretation is only possible because these economists have posited the existence of an 'inner rational agent' endowed with preferences which are consistent and stable over time. But nothing in the field of psychology can justify this assumption. And if there is no 'inner rational agent' to be found somewhere inside the acting individual, inconsistencies are not necessarily mistakes. This flaw can also be found in the representation of the agent underlying the multiple selves model. Someone who would make the New Year's resolution to never again drink alcohol, but would later in the year order a glass of wine at the restaurant contradicts her initial plan. For Thaler and Sunstein, it must be that there is a 'good' and consistent course of action that this person would have followed if only she had 'complete self-control'. The self who is making New Year's resolutions should most likely be identified as the 'planner self' whose interests reflect those of the person. But if we reject the implicit assumption that there *must* be a good and consistent course of action, this conclusion does not follow.

> Both when she was making New Year's resolution and when she was in the restaurant, she had to strike a balance between con-

[9] 'If somebody now wants our help in constraining his later behavior against his own wishes at this later time, how do we decide which side we are on?' (Schelling 1984, 87)

> siderations that pointed out in favour of alcohol and considerations that pointed against it. The simplest explanation of her behaviour is that she struck one balance in the first case and a different balance in the second. This is not a self-control problem; it is a change of mind. (Sugden 2018a, 81)

The fact that we are inclined to categorize this inconsistent behaviour as something which is a 'problem' (of self-control) rather than simply a 'change' (of mind) would be a product of the 'inner rational agent' fallacy, which mislead some economists to believe that a given behaviour is the result of a mistake or a lack of self-control if it is not consistent. On the contrary, the fact that individuals act inconsistently in their daily lives would normally falsify the 'inner rational agent' assumption, but the model of multiple selves

> cannot recognize the continuing identity and agency of ordinary human beings who happen to choose in ways that disconfirm the received theory. A failure of the theory is being re-cast as a failure of the individuals whose behaviour the theory is supposed to explain. (ibid., 105)

Suppose that someone does not save enough during his working life and ends up with a meagre retirement pension or that someone else has lost a lot of money because she paid for a subscription to the gym but has never set foot there. Sugden would say that, at each point in time, this person has done what she wanted at the moment when she wanted it, which does not call for any outside intervention. But as Sugden recognizes, this particular conclusion is warranted because he assumes a 'continuing' agent, which is the same at any point in time. What the continuing agent values, according to Sugden, is just whatever she values over time, whenever she has to make a decision. What is surprising is that Sugden does not try to ground this representation of the agent in empirical evidence, although he and his coauthors attacked the representation of the 'inner rational agent' for lacking psychological foundations. He offers it to his reader as an alternative to the multiple selves model. This suggests that, for Sugden, evidence about human behaviour cannot determine the adoption of a particular representation of the agent. A different representation may lead to different normative judgments and policy recommendations, as we have seen, but the choice of this representation may be fundamentally underdetermined by the evidence

on human behaviour and thus, derives from normative premises that bridge this gap and that one should make explicit.

Indeed, Sugden's conception of the identity of the agent has a normative character. According to Sugden's contractarianism, the role of the economist is to propose to individuals that they take certain actions or undergo certain changes that will generate a collective arrangement which is in everyone's interest. This is only possible if a certain representation of the agents' interests, and also a certain representation of the agents themselves, is provided. Sugden proposes to adopt an opportunity criterion according to which it is in the best interest of everyone to have more opportunity than less. In this framework, an agent is to be conceived as someone 'responsible' who 'treats her past actions as her own, whether or not they were what she now desires them to have been. She treats her future actions as her own, even if she does not know what they will be, and whether or not she expects them to be what she now desires them to be' (Sugden 2018a, 106). A responsible agent may experience regret. But he also values the fact that he has chosen what he wanted when he wanted it. This representation of agents' interests and identity achieves its goal if the individuals who are, according to Sugden, the true addressee of the economist's recommendation, can recognize themselves in it. I will suppose that this is the case, and, in the following, adopt the point of view of Sugden's responsible agents.

What to make of the situation where such a responsible agent is asked to use a HCD? Suppose that she would use it. This would imply that she does not 'treat her future actions as her own', precisely because she expects them not to be 'what she now desires them to be'. She would not be a responsible agent, according to Sugden's characterization[10]. Someone who buys or accepts to use a HCD is revealing that she expects to have a self-control problem (and not a simple change of mind) since she feels the need to limit the actions of her future self. Besides, her own decision does not coincide with Sugden's opportunity criterion, since by closing some of her options, she shows how preferable it is for her to have fewer opportunities rather than more. The criterion that the economist would use to evaluate possible changes would thus clash with the agent's own evaluation of her situation. To sum up, the preference for commitment that individuals reveal when they use a HCD seems to falsify Sugden's representation of the agent, because no responsible individuals, it would appear, would ever choose to

[10]See Fumagalli (2023, 11).

use a HCD. On the contrary, the use of HCDs seems to support the idea that individuals are not responsible, in a way that a model of multiple selves can capture[11].

What makes Sugden's view puzzling is that he clearly denies that using HCD has either a positive or negative value with regard to his opportunity criterion[12], and gives them zero value, which puts his own evaluation criterion on a par with the standard welfare criterion. But contrary to standard welfare economics, which assumes that the choices of individuals are consistent and stable over time, Sugden makes no such assumption. The opportunity criterion is supposed to capture the fact that it is in the interests of individuals to be able to change their minds, provided that they can see themselves as responsible. But if individuals are truly responsible agents, the use of HCDs must be out of the picture, because it is truly incompatible to treat one's future actions, whatever they may be, as one's own and at the same time do everything to prevent them from happening. Sugden's comments about HCDs reflect this confusion:

> If a person knows that she sometimes wants to constrain her future choices, she might reasonably think it in her interest to have certain opportunities for self-constraints. Or, just as reasonably, she might think the opposite. Knowing that, if there are opportunities for self-constraint, she will sometimes find that she is unable to do what she wants because of a constraint that she had previously imposed on herself but now wishes she hadn't, she might think it in her interest that such opportunities are *not* made available. Which view she takes seems to depend on whether, at the time she is making the judgement about her interests, she identifies with the self that imposes the constraint or with the self that is constrained. (Sugden 2018a, 150-151)

Sugden here presents a false equivalence. It is clearly in the interest (in Sugden's sense) of the constrained self to free herself from her previous commitment and act according to her preferences, even if this contradicts the

[11]'Thomas Schelling observed that when people use self-command, to prevent their future selves from acting waywardly, they effectively divide themselves into two selves with conflicting desires for the same point' (Read 2006, 681).

[12]Sugden seems willing to make some exceptions and consider that HCDs are sometimes good for individuals, which makes his position even more puzzling.

plans of her previous self. But if it is also in the interest of the present self to constrain her future choices, then she is not a responsible individual, and she is not the addressee of the contractarian economist's recommendation. A normative economics approach based on Sugden's opportunity criterion would simply not apply to her. As she is willing to pay something to constrain herself, her behaviour reveals that she is faced with what is best described as self-control problems, and not a mere change of mind. In terms of normative evaluation, a completely different approach would be needed to adequately address her interests, one which would not, as Sugden does, exclude a paternalistic intervention[13].

The difficulty Sugden faces seems to derive from the fact that his framework is explicitly presented as a defence of the market. His opportunity criterion should not, therefore, involve preventing the development of markets for HCDs, where individuals engage in transactions that are in their mutual interest—at least at the point in time at which they choose to commit themselves. As we shall see, Mill's position on this issue reflects a similar difficulty, but the argument that Mill constructs, and that Sugden, though inspired by the liberal tradition which stems from Mill, does not mention, shows a possible way out of this difficulty, for those who accept the same normative premises as Mill and Sugden.

5.4 HCDs and Mill's liberty principle

As we have seen, neither the perspective of the multiple selves model nor the perspective of Sugden's responsible agents can make sense of the disturbing character of HCDs. This section will explore a completely different approach, concerned with consistency in the application of a normative principle, rather than with consistency of choice behaviour. The reason why HCDs can be seen as bad as such is not that the agents would be always worse off as a result of using a HCD, or because it is not in their interest as responsible agents, but simply because we cannot, for the sake of individual freedom, allow people to renounce their freedom. According to Mill,

[13]If self-control problems are taken seriously, it is needed, as was remarked earlier, to determine whether the agent is sophisticated enough or naïve, in which case a paternalistic intervention might be warranted because the agent would not choose by herself to use a HCD, or not the right one even if she needs it. The recognition of the reality of self-control problems opens the way to paternalism, which Sugden refuses.

> The reason for not interfering, unless for the sake of others, with a person's voluntary act, is consideration for his liberty. His voluntary choice is evidence that what he so chooses is desirable, or at the least endurable, to him, and his good is on the whole best provided for by allowing him to take his own means of pursuing it. But by selling himself for a slave, he abdicates his liberty; he forgoes any future use of it beyond that single act. He therefore defeats, in his own case, the very purpose which is the justification of allowing him to dispose of himself. He is no longer free; but is thenceforth in a position which has no longer the presumption in its favour, that would be afforded by his voluntary remaining in it. The principle of freedom cannot require that he should be free not to be free. (Mill 1859/2006, 115-116)

This argument has sometimes been seen by commentators, such as Dworkin (1972), as making an exception to Mill's liberty principle (sometimes also called 'harm principle'), according to which 'adults should be free from legal or societal constraints to do what they want to do, provided that their chosen actions do not adversely affect others' (Archard 1990, 453). It would appear that committing oneself to become someone's else slave would not harm anyone else than oneself, and therefore contradict Mill's rule that 'the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others' (Mill 1859/2006, 16). But David Archard convincingly showed that this exception is in fact consistent with the *intention* to implement the liberty principle resolutely.

The reason why the liberty principle is so important to Mill is because it preserves the exercise of individual freedom. Chapters II and III of *on Liberty* have outlined the general (and instrumental) reasons why freedom is valuable: in essence, because it shapes a space for the development of individuality (Sugden 2003). From this point of view, a slavery contract represents a loss of freedom, in that it makes it impossible for the slave to later exercise his freedom, thus restricting the further development of his individuality. But at the same time, the act of committing oneself to become a slave is in itself a valued exercise of freedom by the person who has chosen to do it. It is therefore difficult to decide whether or not it is good, from the perspective of individual freedom, that people are allowed to enter into such a contract: we find ourselves in the same difficulty in which Sugden was.

And yet, as Mill's quoted argument makes clear, the enforcement of a slavery contract would be a self-defeating consequence of the liberty principle, should it allow it. It would make it possible to put an end to the exercise of individual freedom, whereas the protection of this exercise is what the liberty principle was designed for. More precisely, entering into a slavery contract is a case of what Archard calls a 'self-abrogating exercise' of the capacity to choose as one wishes. According to Archard, some action 'is a self-abrogating exercise of y by x if x's doing [it] brings it about that x cannot subsequently exercise y' (Archard 1990, 459). Other examples include voting to abolish elections, using one's freedom of expression to self-censor or to argue to put an end to it, using one's reason or education to alter one's judgement and become a fanatic, etc. All these examples are disturbing because they contradict the obvious reason why voting, freedom of expression, education, etc. were set up in the first place, which is to build and protect the exercise of a valuable capacity. Archard derives from this the general argument that 'where principles are justified by the fact of their guaranteeing something valuable, it is inconsistent with these principles to allow anything which denies or abolished what they seek to guarantee'. Since Mill's liberty principle is justified because it guarantees the exercise of individual freedom (subject to the harm condition), 'it would be inconsistent with holding that principle justified to permit behaviour which denied the exercise of freedom'.

It is clear, however, that just as freedom of expression laws cannot as such forbid self-censorship, which is a form of expression, the liberty principle cannot forbid people to commit themselves to obey someone else. But the previous argument gives a reason to refuse to enforce the contract, in the case where the slave would change his mind and renege on his commitment. Because Archard's reformulation of Mill's argument gives a very general form to it, it would apply to any 'self-abrogating exercise' of a valuable capacity, which corresponds exactly to what HCDs are. A HCD prevents someone from doing something which he knows would be valuable for himself at another point in time. In the absence of regulations, a market for HCDs would have the self-defeating consequence, with regard to the liberty principle, that people may lose their capacity to enjoy their freedom to do certain things for any possible length of time. A slavery contract can be seen as a dramatic extension of this mechanism. Note that this argument only concerns hard commitment devices, because only arrangements related to HCDs would need to be enforced by an outside agent, such as the state. If someone could refuse to pay the amount of money he committed to pay should he fail to reach a

given target, we would not speak of a hard commitment device.

An objection often raised to Mill's reasoning is that this argument would prove too much[14]. It would make it impossible to give up certain freedoms or opportunities, which is often necessary to live a decent social life: getting married, or having a job, involves losing much of one's freedom and valuable opportunities. Someone getting married or getting a job commits themselves to losing some opportunities to secure something else with the help of someone else: stable income, lasting love, etc. Crucially, this commitment is also valuable for others: the other party of the contract, the person who would benefit from the promise made by someone have something to lose—valuable opportunities—if the contract or the promise cannot be made and enforced. The argument developed here says nothing about voluntarily losing one's freedom to improve others' welfare and opportunities—it only concerns losing one's freedom to ensure that one's actions are consistent with the initial plan that one had about one's self-regarding conduct. It does not seem that the purpose of the liberty principle is defeated when someone's loss of freedom is meant to enhance other people's freedom. But if this consequence fails to materialize, the principle may be defeated. Let us therefore define a 'pure' self-abrogating exercise of some capacity as one that can *only* be done for the sake of rendering impossible the exercise of one's own capacity. The act of using a HCD is a pure self-abrogating exercise of freedom since it is done with the intention to give up one's freedom to solve a self-control problem, which is achieved by imposing consistency on one's actions without benefiting directly anyone else. The argument would thus object to the existence of enforceable 'hard' commitment contracts, but not to other enforceable contracts.

Conclusion: self-nudging and the capacity for self-control

The last section made the case that using HCDs is bad as such because it is a self-abrogating exercise of individual freedom, and that markets for HCDs should be regulated so that individuals are not forced against their will to incur material penalties as they are supposed to if they fail to deliver on their commitment. HCDs can also be bad for other reasons, related to

[14]See in particular Lovett (2008, 130-132).

welfare losses, which were detailed in section 2. That being said, I am not denying that self-control problems are real, and my point is not that sane adults who may incur significant losses due to these problems should just swallow the pill and take it as the responsible individuals they should be[15]. What Schelling said in 1985 seems to be just as true today, if not more: 'by and large, people are more in need of greater efficacy in devising rules of their own than in danger of shortsighted self-binding activity'. Reijula and Hertwig (2022) agree: 'past and existing levels of self-control no longer suffice to enable self-governance in these finely tuned choice environments' that make the most of cognitive bias and temptations to nudge consumers into buying and consuming goods that they may not have bought otherwise. But a market for HCDs cannot be the answer, as HCDs are bad as such and, according to what I called the incentives argument, they may discourage people from building their own capacity for self-control and instead rely on products supplied by private firms which may exploit them[16].

A valuable—and fully compatible with individual freedom—alternative to a market for HCDs is the practice of self-nudging, as Reijula and Hertwig (2022) describe it. Georges Ainslie's work, in particular, has cast a light on the ways individuals may practice self-management without needing a hard commitment. Most of the various self-nudging practices described by Reijula and Hertwig—which they defined as 'tools for promoting self-knowledge and internal negotiation between the various needs and desires inhabiting people's minds and bodies'—correspond to the use of soft commitment devices to solve self-control problems. A psychological cost (such as shame, 'frictions', etc.) is attached to certain options by the self-nudging practices, which prevents ulterior selves from choosing them. This requires individuals to be active in their self-management, and self-aware of their own biases, temptations, and more generally their own psychology, which is exactly why self-nudging is a good way to address the incentive argument. If individuals cannot rely on a HCD to solve their self-control problems, they are encouraged both to address the problems themselves without risking incurring penalties and to invest in self-awareness and mastery of self-management techniques. Rationality is thus somehow restored, because 'rational agency is sometimes approximated

[15]That does not seem to be Sugden's opinion either, but as I tried to show, it is not clear why this should not be his conclusion.

[16]See in particular the problem raised by 'partially naive' agents who do not commit enough (Eliaz and Spiegler 2006; Della Vigna and Malmendier 2004, 2006), and who thus can be exploited.

thanks to good habits, rules and scaffolding institutions' (Reijula and Hertwig 2022, 136). The main takeaway of these approaches is that the fact that the agent is rational or a 'continuous locus of responsibility' (Sugden) is not something to be assumed or rejected, but something that we can (and should) *make happen*.

Public interventions may be useful to achieve this because they can promote practices of self-nudging and make individuals aware of the extent of their own self-control problem. Besides, as emphasized by Reijula and Hertwig, self-nudges eschew the major ethical and practical criticisms that are often addressed against paternalistic nudges: impairment of autonomy, difficulty of preference identification, unintended side effects, etc. None of this is really new: in the absence of markets for HCDs, people have always developed more or less elaborated techniques to overcome their self-control problems. The attempt to incorporate scientific behavioural evidence in the practice of self-management is not new either: Descartes's classical essay *The Passions of the Soul* (1649/2015) is a prominent example of that. But Reijula and Hertwig's call for individuals to 'take back power' by taking advantage of the psychological and behavioural insights that are often used to nudge them unwittingly is fully in line with the liberal tradition of John Stuart Mill and its promotion of the value of 'individuality', while not assuming away the existence of self-control problems.

General Conclusion

Anti-paternalism and values

The primary task of normative economics is to evaluate policies and institutions, based on certain values. In welfare economics, which is the main form normative economics took in the 20th century, the key criterion for these evaluations is preference satisfaction. Within this framework, anti-paternalism—the idea that it is for each individual to judge what matters for her—serves two distinct roles:

- When it comes to the justification of paternalistic *interventions*, anti-paternalism is seen as *implied* by the traditional assumptions about the rationality of the economic agent. If individuals are presumed to know what is best for them, paternalistic interventions that interfere with their choices are suboptimal because they could only worsen the situations for these agents, making such interventions undesirable.

- When it comes to *evaluations*, 'normative minimalism' (Haybron and Alexandrova 2013), endorsed by many economists—the attempt to 'keep value commitments to a minimum' involves deferring to individuals' own preferences and values. Anti-paternalism in this sense is a matter of principle, not a consequence.

With the advent of behavioural economics, the first form of anti-paternalism described above has been undermined. If individuals cannot consistently make choices that serve their best interests, paternalistic intervention may become justifiable. However, the second form of anti-paternalism, which emphasizes respecting individual preferences and values, continues to be appealing to economists. Behavioural paternalists, such as Sunstein and

Thaler, support the 'means paternalism' program, which aims to assist individuals in achieving their objectives. 'Normative minimalism' would not allow economists to substitute their own ends for those of individuals. Therefore, an intervention that interferes with individuals' choices only to correct their mistakes or failures to reach their goal may not depart too much from the second form of anti-paternalism I described. However, it raises the question of how to identify individuals' ends, a topic not covered in this thesis.

This thesis is about freedom. When considering a paternalistic intervention, one may always question whether it could have been avoided by allowing individuals to act on their own. Even if interventions like nudging can enhance individuals' prospects, one may ask whether individuals can be empowered to self-nudge instead of imposing nudges upon them, to the same effect (Reijula and Hertwig 2022). A limitation of 'normative minimalism' and the preference satisfaction criterion is their inability to answer whether, when the outcomes are the same, it is preferable for public authorities to intervene and restrict individuals' choices, or empower individuals to protect themselves from failing to reach their goals, or let the market provide such devices. To address this question, there is a need to make more room for judgements about freedom and commitments. The thesis developed several approaches to enrich the normative economist's toolbox and meet this need.

Interventions and freedom

The thesis explored two types of interventions through which public authorities interfere with individuals' choices:

- *Interventions to protect individuals from themselves*: These interventions aim to prevent individuals from failing to achieve their goals, as discussed in chapters 3 and 5.

- *Interventions to solve collective problems*: These interventions address issues arising from externalities or public goods on a societal scale, as discussed in chapter 2.

In both cases, the central question is whether such interventions align with a concern for freedom. If individuals can be provided with commitment devices or binding contracts to help them make choices, or if markets can handle these functions, why should public authorities interfere with their

choices? A concern for freedom may seem to imply a presumption in favour of *laissez-faire*. As we have seen, a perspective akin to Nozick's negative freedom would exclude coercive interventions and favour the private provision of commitment devices and suitable contracts. However, the thesis considers scenarios where this is not possible or desirable:

- In what I called 'situations of urgency' (chapter 2), setting up assurance contracts to resolve public good problems may be infeasible. In such situations, an intervention may be justified from the standpoint of 'extended libertarianism,' which integrates concerns for 'indirect liberty' alongside the familiar idea of 'liberty as control.' Based on Amartya Sen's insight that it may be relevant to speak of freedom even if individuals have no control over their situation, extended libertarianism permits coercive interventions, provided they only impose what individuals would willingly impose on themselves.

- As demonstrated by Schelling, in strategic interactions (chapter 3), people are sometimes made vulnerable simply by having choices when confronted by opponents who can inflict harm. In such cases, offering a choice not to choose (a commitment device) may be ineffective, making 'strategic paternalism' relevant. Strategic paternalism corrects individuals' failures to reach their goals by establishing the credibility of their contingent plans and altering others' expectations, ultimately furthering individuals' ends. It accomplishes the means paternalism program by taking as given people's ends and helping to reach them.

The thesis introduces two conceptual innovations, 'extended libertarianism' and 'strategic paternalism'. In both cases, the interventions steer individuals toward outcomes they would achieve in ideal circumstances if they had the opportunity to commit or bind themselves. In this regard, the two classes of interventions are justifiable from the perspective of freedom, or more precisely, 'indirect liberty', if certain conditions are satisfied. The thesis concludes that a concern for freedom, even in its negative form, does not necessarily imply a presumption for *laissez-faire*. It is compatible with constraints (even severe constraints), self-imposed or not, if we accept that the domains of freedom and control do not overlap. Commitments play a crucial role because they reveal the price individuals are willing to pay in terms of relinquishing some control to attain the goals they hold dear. They provide a basis for evaluating policies and institutions in terms of freedom, enabling

us to understand and respect the trade-offs individuals make between being in control and the realization of their goals. This understanding helps in assessing the extent to which individuals are ready to accept or reject the loss of control inherent in various policies or institutions.

List of Tables

Bibliography

Ahlert, M. (2010). A New Approach to Procedural Freedom in Game Forms. *Special Issue on Ethics and Economics 26*(3), 392–402.

Ainslie, G. (1992). *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person*. Cambridge: Cambridge University Press.

Alexandrova, A. (2018). Can the Science of Well-Being Be Objective? *The British Journal for the Philosophy of Science 69*(2), 421–445.

Anderson, E. (1990). The Ethical Limitations of The Market. *Economics & Philosophy 6*(2), 179–205.

Anderson, E. (2001). Symposium on Amartya Sen's Philosophy: 2 Unstrapping the Straitjacket of 'preference': a Comment on Amartya Sen's Contributions to Philosophy and Economics. *Economics & Philosophy 17*(1), 21–38.

Andreoni, J. (1990). Impure Altruism and Donations to Public Good. *The economic journal 100*(401), 464–477.

Andreoni, J. (2006). Philanthropy. In S.-C. Kolm and J. M. Ythier (Eds.), *Handbook of the Economics of Giving, Altruism and Reciprocity*, Volume 2, pp. 1201–1269. Amsterdam: Elsevier.

Andreou, C. (2022). *Commitment and Resoluteness in Rational Choice*. Cambridge: Cambridge University Press.

Angner, E. (2019). We're All Behavioral Economists Now. *Journal of Economic Methodology 26*(3), 195–207.

Archard, D. (1990). Freedom Not to Be Free: the Case of the Slavery Contract in JS Mill's On Liberty. *The Philosophical Quarterly 40*(161), 453–465.

Archard, D. (2002). Selling Yourself: Titmuss's Argument Against a Market in Blood. *The Journal of Ethics 6*, 87–102.

Arrow, K. J. (1972). Gifts and Exchanges. *Philosophy & Public Affairs 1*(4), 343–362.

Arrow, K. J. (2012). *Social Choice and Individual Values*. New Haven: Yale University Press.

Atkinson, A. B. (2001). The Strange Disappearance of Welfare Economics. *Kyklos 54*(2-3), 193–206.

Bader, R. M. (2018). Moralized Conceptions of Liberty. In D. Schmidtz and C. E. Pavel (Eds.), *The Oxford Handbook of Freedom*. New York: Oxford University Press.

Badiei, S., G. Campagnolo, and A. Grivaux (2022). *Le Positif, le normatif et la philosophie économique*. E-conomiques. Paris: Éditions Matériologiques.

Banerjee, A. and E. Duflo (2019). *Good Economics for Hard Times: Better Answers to Our Biggest Problems*. United Kingdom: Penguin Books.

Banerjee, A. V. and S. Mullainathan (2008). Limited Attention and Income Distribution. *American Economic Review 98*(2), 489–493.

Barberà, S., W. Bossert, and P. K. Pattanaik (2004). Ranking Sets of Objects. In S. Barberà, P. Hammond, and C. Seidl (Eds.), *Handbook of Utility Theory. Volume 2, Extensions*, pp. 893–977. Boston: Springer.

Baujard, A. (2007). Conceptions of Freedom and Ranking Opportunity Sets. A Typology. *Homo Oeconomicus 24*(2), 1–24.

Baujard, A. (2011). Utilité et liberté de choix dans les classements d'ensembles d'opportunités. *Raisons Politiques 43*(03), 59–92.

Baujard, A. (2016). Welfare Economics. In G. Faccarello and H. D. Kurz (Eds.), *Handbook on the History of Economic Analysis Volume III: Developments in Major Fields of Economics*, Handbook on the History of Economic Analysis. Cheltenham: Edward Elgar Publishing.

Baujard, A. (2017). L'économie du bien-être est morte. Vive l'économie du bien-être ! In *Philosophie économique*, E-conomiques, pp. 77–128. Paris: Éditions Matériologiques.

Beauchamp, T. L. (2005). The Nature of Applied Ethics. In R. G. Frey and C. Wellman (Eds.), *A Companion to Applied Ethics*, pp. 1–16. Hoboken: John Wiley and Sons.

Beck, L. and M. Jahn (2021). Normative Models and Their Success. *Philosophy of the Social Sciences 51*(2), 123–150.

Becker, G. S. and K. M. Murphy (1988). A Theory of Rational Addiction. *Journal of political Economy 96*(4), 675–700.

Berlin, I. (2002). *Liberty*. Oxford: Oxford University Press.

Bernheim, B. D. (2009). Behavioral Welfare Economics. *Journal of the European Economic Association 7*(2-3), 267–319.

Bernheim, B. D. and A. Rangel (2009). Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. *The Quarterly Journal of Economics 124*(1), 51–104.

Bervoets, S. (2007). Freedom of Choice in a Social Context: Comparing Game Forms. *Social Choice and Welfare 29*(2), 295–315.

Binder, C. (2019). *Agency, Freedom and Choice*. Dordrecht: Springer Netherlands.

Binder, C. (2021a). Beyond Welfarism. In R. E. Backhouse, A. Baujard, and T. Nishizawa (Eds.), *Welfare Theory, Public Action, and Ethical Values: Revisiting the History of Welfare Economics*, pp. 277–297. Cambridge: Cambridge University Press.

Binder, C. (2021b). Freedom and Markets. In J. Reiss and C. Heilmann (Eds.), *The Routledge Handbook of Philosophy of Economics*, pp. 457–466. New York: Routledge.

Binder, C. and I. Robeyns (2019). Economic Ethics and the Capability Approach. In M. D. White (Ed.), *The Oxford Handbook of Ethics and Economics*. Oxford: Oxford University Press.

Blaug, M. (1992). *The Methodology of Economics: Or, How Economists Explain*. Cambridge Surveys of Economic Literature. Cambridge: Cambridge University Press.

Bowles, S. (2017). *The Moral Economy: Why Good Incentives are No Substitute for Good Citizens*. New Haven: Yale University Press.

Braham, M. (2006). Measuring Specific Freedom. *Economics & Philosophy 22*(3), 317–333.

Braham, M. and M. van Hees (2014). The Impossibility of Pure Libertarianism. *The Journal of Philosophy 111*(8), 420–436.

Brennan, J. and P. Jaworski (2016). *Markets Without Limits: Moral Virtues and Commercial Interests*. New York: Routledge.

Brink, D. O. (2013). *Mill's Progressive Principles*. Oxford: Oxford University Press.

Bryan, G., D. Karlan, and S. Nelson (2010). Commitment Devices. *Annual Review of Economics 2*(1), 671–698.

Buchanan, J. M. (1999). *The Logical Foundations of Constitutional Liberty*. Indianapolis: Liberty Fund.

Camerer, C., S. Issacharoff, G. Loewenstein, T. O'donoghue, and M. Rabin (2003). Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism". *University of Pennsylvania law review 151*(3), 1211–1254.

Carter, I. (1995). The Independent Value of Freedom. *Ethics 105*(4), 819–845.

Carter, I. (1999). *A Measure of Freedom*. New York: Oxford University Press.

Carter, I. (2004). Choice, Freedom, and Freedom of Choice. *Social Choice and Welfare 22*(1), 61–81.

Carter, I. (2011). Distributing Freedom over Whole Lives. In A. Gosseries and P. Vanderborght (Eds.), *Arguing about Justice: Essays for Philippe Van Parijs*. Louvain-la-Neuve: Presses universitaires de Louvain.

Carter, I. (2022). Positive and Negative Liberty. https://plato.stanford.edu/archives/spr2022/entries/liberty-positive-negative/. Accessed on September 2023.

Cohen, G. A. (2011). *On the Currency of Egalitarian Justice, and Other Essays in Political Philosophy*. Princeton: Princeton University Press.

Conly, S. (2013). *Against Autonomy: Justifying Coercive Paternalism*. Cambridge: Cambridge University Press.

Costella, A. (2023). Adaptive Preferences, Self-Expression and Preference-Based Freedom Rankings. *Economics & Philosophy*, 1–22.

Courtwright, D. T. (2019). *The Age of Addiction: How Bad Habits Became Big Business*. Cambridge, MA: The Belknap Press of Harvard University Press.

Cowen, T. (1991). Self-Liberation Versus Self-Constraint. *Ethics 101*, 360.

DellaVigna, S. and U. Malmendier (2004). Contract Design and Self-Control: Theory and Evidence. *The Quarterly Journal of Economics 119*(2), 353–402.

DellaVigna, S. and U. Malmendier (2006). Paying Not to Go to the Gym. *American Economic Review 96*(3), 694–719.

Descartes, R. (2015). *The Passions of the Soul and Other Late Philosophical Writings*. Oxford: Oxford University Press.

Desmarais-Tremblay, M. (2017). Musgrave, Samuelson, and the crystallization of the standard rationale for public goods. *History of Political Economy 49*(1), 59–92.

Desmarais-Tremblay, M. (2020). W.H. Hutt and the conceptualization of consumers' sovereignty. *Oxford Economic Papers 72*(4), 1050–1071.

Dietrich, F. and C. List (2016). Mentalism versus behaviourism in economics: a philosophy-of-science perspective. *Economics & Philosophy 32*(2), 249–281.

Dixit, A. K. and B. Nalebuff (2008). *The Art of Strategy: a Game Theorist's Guide to Success in Business and Life*. New York: WW Norton & Company.

Dold, M. and P. Lewis (2023). A Neglected Topos in Behavioural Normative Economics: The Opportunity and Process Aspect of Freedom. *Behavioural Public Policy 7*(4), 943–953.

Dowding, K. and M. van Hees (2007). Counterfactual Success and Negative Freedom. *Economics and Philosophy 23*(2), 141–162.

Dowding, K. and M. van Hees (2009). Freedom of Choice. In P. Anand, P. Pattanaik, and C. Puppe (Eds.), *Handbook of Rational and Social Choice*. Oxford: Oxford University Press.

Duflo, E. (2012). Tanner Lectures on Human Values and the Design of the Fight Against Poverty. *Manuscript, MIT*.

Dupré, J. (2007). Fact and Value. In H. Kincaid, J. Dupré, and A. Wylie (Eds.), *Value-Free Science? Ideals and Illusions*. New York: Oxford University Press.

Dworkin, G. (1972). Paternalism. *the Monist 56*(1), 64–84.

Dworkin, G. (1982). Is More Choice Better Than Less? *Midwest studies in Philosophy 7*(1), 47–61.

Dworkin, R. (1985). *A Matter of Principle*. Cambridge, MA: Harvard University Press.

Eliaz, K. and R. Spiegler (2006). Contracting with Diversely Naive Agents. *The Review of Economic Studies 73*(3), 689–714.

Elliott, K. C. (2022). *Values in Science*. Elements in the Philosophy of Science. Cambridge: Cambridge University Press.

Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge paperback library. Cambidge: Cambridge University Press.

Elster, J. (1984). *Ulysses and the Sirens: Studies In Rationality And Irrationality*. Cambridge: Cambridge University Press.

Elster, J. (2000). *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge: Cambridge University Press.

Ergin, H. and T. Sarver (2010). A Unique Costly Contemplation Representation. *Econometrica 78*(4), 1285–1339.

Favereau, J. (2021). *Le Hasard de la preuve: Apports et limites de l'économie expérimentale du développement*. Lyon: ENS Éditions.

Fleurbaey, M. (1996). *Théories économiques de la justice*. Paris: Economica.

Fleurbaey, M. (2006). Capabilities, Functionings and Refined Functionings. *Journal of Human Development 7*(3), 299–310.

Fleurbaey, M. (2012). The Importance of What People Care About. *Politics, Philosophy & Economics 11*(4), 415–447.

Fleurbaey, M. (2023). Normative Economics and Economic Justice. `https://plato.stanford.edu/archives/fall2023/entries/economic-justice/`. Accessed on September 2023.

Fontaine, P. (2002). Blood, Politics, and Social Science: Richard Titmuss and the Institute of Economic Affairs, 1957–1973. *Isis 93*(3), 401–434.

Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of the Emotions.* New York: WW Norton & Company.

Frey, B. S. and F. Oberholzer-Gee (1997). The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding- Out. *The American Economic Review 87*(4), 746–755.

Frey, B. S. and A. Stutzer (2002). What Can Economists Learn From Happiness Research? *Journal of Economic literature 40*(2), 402–435.

Friedman, M. (1953). The Methodology of Positive Economics. In *Essays in Positive Economics*, pp. 3–43. Chicago: The University of Chicago Press.

Friedman, M. (2002). *Capitalism and Freedom*. Chicago: University of Chicago press.

Friedman, M. and R. D. Friedman (1990). *Free to Choose: a Personal Statement*. San Diego: Mariner Books Classics.

Fudenberg, D. and J. Tirole (1991). *Game theory*. Cambridge, MA: MIT press.

Fumagalli, R. (2023). Preferences Versus Opportunities: On the Conceptual Foundations of Normative Welfare Economics. *Economics & Philosophy*, 1–25.

Gaertner, W., P. K. Pattanaik, and K. Suzumura (1992). Individual Rights Revisited. *Economica 59*(234), 161–177.

Gibbard, A. (1974). A Pareto-Consistent Libertarian Claim. *Journal of Economic Theory 7*(4), 388–410.

Gneezy, U., S. Meier, and P. Rey-Biel (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives 25*(4), 191–210.

Gneezy, U. and A. Rustichini (2000). A Fine Is a Price. *The Journal of Legal Studies 29*(1), 1–17.

Goodin, R. E. (1986). Laundering Preferences. *Foundations of social choice theory 75*, 81–86.

Gray, T. (1990). *Freedom*. London: Palgrave Macmillan.

Grüne-Yanoff, T. (2012). Old Wine in New Casks: Libertarian Paternalism Still Violates Liberal Principles. *Social Choice and Welfare 38*(4), 635–645.

Gul, F. and W. Pesendorfer (2001). Temptation and Self-Control. *Econometrica 69*(6), 1403–1435.

Gärdenfors, P. (1981). Rights, Games and Social Choice. *Noûs 15*(3), 341–356.

Hands, D. W. (2008). Philosophy and Economics. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics: Volume 1 – 8*, pp. 4922–4932. London: Palgrave Macmillan UK.

Hart, H. L. A. (1973). Rawls on Liberty and Its Priority. *The University of Chicago Law Review 40*(3), 534–555.

Hausman, D. M. (2011). *Preference, Value, Choice, and Welfare*. New York: Cambridge University Press.

Hausman, D. M. (2018). Behavioural Economics and Paternalism. *Economics & Philosophy 34*(1), 53–66.

Hausman, D. M. (2021). Philosophy of economics: past and future. *Journal of Economic Methodology 28*(1), 14–22.

Hausman, D. M., M. McPherson, and D. Satz (2016). *Economic Analysis, Moral Philosophy, and Public Policy*. New York: Cambridge University Press.

Hausman, D. M. and B. Welch (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy 18*(1), 123–136.

Haybron, D. M. and A. Alexandrova (2013). Paternalism in economics. In C. Coons and M. E. Weber (Eds.), *Paternalism: Theory and practice*, pp. 157–177. Cambridge: Cambridge University Press.

Hayek, F. A. (2011). *The Constitution of Liberty: The Definitive Edition*. The Collected Works of F. A. Hayek. Chicago: University of Chicago Press.

Herrnstein, R. J., G. F. Loewenstein, D. Prelec, and W. Vaughan Jr. (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioral Decision Making 6*(3), 149–185. Publisher: John Wiley & Sons, Ltd.

Herzog, L. (2021). Markets. `https://plato.stanford.edu/archives/fall2021/entries/markets/`. Accessed on September 2023.

Hutt, W. H. (1940). The Concept of Consumers' Sovereignty. *The Economic Journal 50*(197), 66–77.

Infante, G., G. Lecouteux, and R. Sugden (2016). Preference Purification and the Inner Rational Agent: A Critique of the Conventional Wisdom of Behavioural Welfare Economics. *Journal of Economic Methodology 23*(1), 1–25.

Iyengar, S. S. and M. R. Lepper (2000). When Choice Is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of personality and social psychology 79*(6), 995.

Jones, P. and R. Sugden (1982). Evaluating Choice. *International review of Law and Economics 2*(1), 47–65.

Kahneman, D. (2011). *Thinking, Fast and Slow*. United Kingdom: Penguin Books.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives 5*(1), 193–206.

Kahneman, D., P. P. Wakker, and R. Sarin (1997). Back to Bentham? Explorations of experienced utility. *The quarterly journal of economics 112*(2), 375–406.

Keynes, J. M. (2010). *Essays in Persuasion*. Basingstoke: Palgrave Macmillan.

Kolm, S.-C. (1985). *Le Contrat social libéral: philosophie et pratique du libéralisme*. Paris: PUF.

Kreps, D. M. (1979). A Representation Theorem for "Preference for Flexibility". *Econometrica 47*(3), 565–577.

Kreps, D. M. (1988). *Notes on the Theory of Choice*. Westview press.

Kreps, D. M. (1990). *Game theory and economic modelling*. Oxford: Clarendon Press.

Kymlicka, W. (2002). *Contemporary Political Philosophy: An Introduction*. Oxford: Oxford University Press.

Le Grand, J. (2007). *The Other Invisible Hand: Delivering Public Services through Choice and Competition*. Princeton: Princeton University Press.

Le Grand, J. (2011). Quasi-Market versus State Provision of Public Services: Some Ethical Considerations. *Public Reason 3*(2), 80–89.

Le Grand, J. and B. New (2015). *Government Paternalism*. Princeton: Princeton University Press.

Le Lec, F. and B. Tarroux (2020). On Attitudes to Choice: Some Experimental Evidence on Choice Aversion. *Journal of the European Economic Association 18*(5), 2108–2134.

Levy, J. T. (2017). Toward a Non-Lockean Libertarianism. In J. Brennan, B. van der Vossen, and D. Schmidtz (Eds.), *The Routledge Handbook of Libertarianism*, pp. 22–33. New York: Routledge.

Lewis, P. and M. Dold (2020). James Buchanan on the Nature of Choice: Ontology, Artifactual Man and the Constitutional Moment in Political Economy. *Cambridge Journal of Economics 44*(5), 1159–1179.

Locke, J. (2016). *Second Treatise of Government and A Letter Concerning Toleration*. Oxford: Oxford University Press.

Lomasky, L. E. (1983). Gift Relations, Sexual Relations and Freedom. *The Philosophical Quarterly (1950-) 33*(132), 250–258.

Lovett, F. (2009). Mill on consensual domination. In C. L. Ten (Ed.), *Mill's On Liberty: A Critical Guide*, Cambridge Critical Guides, pp. 123–137. Cambridge: Cambridge University Press.

MacCallum, G. C. (1967). Negative and Positive Freedom. *The philosophical review 76*(3), 312–334.

Marshall, A. (2013). *Principles of economics* (8th ed.). London: Palgrave Macmillan.

Mauss, M. (2023). *Essai sur le don: Forme et raison de l'échange dans les sociétés archaïques*. Paris: PUF.

Małecka, M. (2021). Values in Economics: A Recent Revival With a Twist. *Journal of Economic Methodology 28*(1), 88–97.

McQuillin, B. and R. Sugden (2012). Reconciling Normative and Behavioural Economics: The Problems to Be Solved. *Social Choice and Welfare 38*(4), 553–567.

Mill, J. and W. Ashley (1909). *Principles of Political Economy with Some of Their Applications to Social Philosophy*. London: Longmans, Green.

Mill, J. S. (2006). *On Liberty and The Subjection of Women*. London: Penguin Classics.

Miller, D. (1983). Constraints on Freedom. *Ethics 94*(1), 66–86.

Mongin, P. and M. Cozic (2018). Rethinking Nudge: Not One but Three Concepts. *Behavioural Public Policy 2*(1), 107–124.

Nozick, R. (1974). *Anarchy, state, and utopia*. New York: Basic Books.

O'Donoghue, T. and M. Rabin (2003). Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes. *American Economic Review 93*(2), 186–191.

O'Neill, O. (2015). IV*—The Most Extensive Liberty. *Proceedings of the Aristotelian Society 80*(1), 45–60.

Oppenheim, F. E. (2004). Social freedom: Definition, measurability, valuation. *Social Choice and Welfare 22*, 175–185.

Ortoleva, P. (2013). The Price of Flexibility: Towards a Theory of Thinking Aversion. *Journal of Economic Theory 148*(3), 903–934.

Page, L. (2022). *Optimally Irrational: The Good Reasons We Behave the Way We Do*. Cambridge: Cambridge University Press.

Pattanaik, P. K. and Y. Xu (1990). On Ranking Opportunity Sets in Terms of Freedom of Choice. *Recherches Économiques de Louvain/Louvain Economic Review 56*(3-4), 383–390.

Pattanaik, P. K. and Y. Xu (1998). On Preference and Freedom. *Theory and Decision 44*, 173–198.

Pattanaik, P. K. and Y. Xu (2009). Individual rights and freedom in welfare economics. In R. Gotoh and P. Dumouchel (Eds.), *Against Injustice: the New Economics of Amartya Sen*. Cambridge: Cambridge University Press.

Pattanaik, P. K. and Y. Xu (2015). Freedom and its Value. In I. Hirose and J. Olson (Eds.), *The Oxford Handbook of Value Theory*, pp. 356–380. New York: Oxford University Press.

Pattanaik, P. K. and Y. Xu (2018). On a concept of freedom. In A. Mishra and T. Ray (Eds.), *Markets, Governance, and Institutions in the Process of Economic Development*. Oxford: Oxford University Press.

Pattanaik, P. K. and Y. Xu (2020). On Capability and its Measurement. In E. Chiappero-Martinetti, S. Osmani, and M. Qizilbash (Eds.), *The Cambridge Handbook of the Capability Approach*, pp. 271–292. Cambridge University Press.

Peil, J. and I. van Staveren (2009). *Handbook of Economics and Ethics*. Elgar Original Reference Series. Cheltenham: Edward Elgar.

Peter, F. (2004). Choice, Consent, and the Legitimacy of Market Transactions. *Economics & Philosophy 20*(1), 1–18.

Prendergast, R. (2005). The Concept of Freedom and Its Relation to Economic Development—a Critical Appreciation of the Work of Amartya Sen. *Cambridge Journal of Economics 29*(6), 1145–1170.

Prendergast, R. (2011). Sen and Commons on Markets and Freedom. *New Political Economy 16*(2), 207–222.

Puppe, C. (1996). An Axiomatic Approach to "Preference for Freedom of Choice". *Journal of Economic Theory 68*(1), 174–199.

Putnam, H. (2002). *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, MA: Harvard University Press.

Radin, M. J. (2001). *Contested Commodities*. Cambridge, MA: Harvard University Press.

Rawls, J. (2005). *A Theory of Justice: Original Edition*. Oxford Paperbacks 301 301. Cambridge, MA: Harvard University Press.

Raz, J. (1988). *The Morality of Freedom*. Oxford: Clarendon Press.

Read, D. (2006). Which Side Are You On? The Ethics of Self-Command. *Journal of Economic Psychology 27*(5), 681–693.

Reijula, S. and R. Hertwig (2022). Self-nudging and the citizen choice architect. *Behavioural Public Policy 6*(1), 119–149.

Reiss, J. (2013). *Philosophy of Economics: A Contemporary Introduction*. Routledge Contemporary Introductions to Philosophy. New York: Routledge.

Rizzo, M. and M. R. Dolde (2020). Can a Contractarian Be a Paternalist? The Logic of James M. Buchanan's System. *Public Choice 183*(3), 495–507.

Rizzo, M. J. and G. Whitman (2020). *Escaping paternalism: Rationality, behavioral economics, and public policy*. Cambridge: Cambridge University Press.

Robbins, L. (1984). *An Essay on the Nature and Significance of Economic Science*. London: Macmillan.

Robeyns, I. (2017). *Wellbeing, Freedom and Social Justice: The Capability Approach Re-examined*. Cambridge: Open Book Publishers.

Rosenberg, A. (1976). *Microeconomic Laws: A Philosophical Analysis*. Pittsburgh: University of Pittsburgh Press.

Roth, A. E. (2007). Repugnance as a Constraint on Markets. *Journal of Economic perspectives 21*(3), 37–58.

Saint-Paul, G. (2011). *The Tyranny of Utility: Behavioral Social Science and the Rise of Paternalism*. Princeton: Princeton University Press.

Samuelson, P. (1972). Indeterminacy of government role in public-good theory. In *The Collected Scientific Papers of Paul A. Samuelson, volume 3*. Cambridge, MA: MIT Press.

Sandel, M. J. (2012). *What Money Can't Buy: The Moral Limits of Markets*. London: Penguin Books.

Sarver, T. (2008). Anticipating Regret: Why Fewer Options May Be Better. *Econometrica 76*(2), 263–305.

Satz, D. (2010). *Why Some Things Should Not Be for Sale: The Moral Limits of Markets*. Oxford Political Philosophy. New York: Oxford University Press.

Scanlon, T. M. (2000). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Schelling, T. C. (1978). Egonomics, or the Art of Self-Management. *The American Economic Review 68*(2), 290–294.

Schelling, T. C. (1980). *The Strategy of Conflict: With a New Preface by the Author*. Cambridge, MA: Harvard university press.

Schelling, T. C. (1984). *Choice and Consequence*. Cambridge, MA: Harvard University Press.

Schilbach, F., H. Schofield, and S. Mullainathan (2016). The Psychological Lives of the Poor. *American Economic Review 106*(5), 435–440.

Schmidt, A. T. (2017). An Unresolved Problem: Freedom across Lifetimes. *Philosophical Studies 174*, 1413–1438.

Schmidtz, D. (1991). *The Limits Of Government: An Essay On The Public Goods Argument*. Boulder: Westview Press.

Schubert, C. (2015). Opportunity and Preference Learning. *Economics & Philosophy 31*(2), 275–295.

Schwartz, B. (2016). *The Paradox of Choice: Why More Is Less, Revised Edition*. New York: HarperCollins.

Searle, J. (1964). How to Derive 'Ought'from 'Is'. *Philosophical Review 73*(1), 43–58.

Sen, A. (1970). The Impossibility of a Paretian Liberal. *Journal of political economy 78*(1), 152–157.

Sen, A. (1977). Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy & Public Affairs 6*(4), 317–344.

Sen, A. (1979). Personal Utilities and Public Judgements: Or What's Wrong with Welfare Economics. *The Economic Journal 89*(335), 537–558.

Sen, A. (1982). Liberty as Control: An Appraisal. *Midwest Studies in Philosophy 7*, 207–221.

Sen, A. (1985). The Moral Standing of the Market. *Social philosophy and policy 2*(2), 1–19.

Sen, A. (1991). Welfare, Preference and Freedom. *Journal of econometrics 50*(1-2), 15–29.

Sen, A. (1993). Markets and Freedoms: Achievements and Limitations of the Market Mechanism in Promoting Individual Freedoms. *Oxford Economic Papers 45*(4), 519–541.

Sen, A. (1994). Markets and the Freedom to Choose. In H. Siebert (Ed.), *The Ethical Foundations of the Market Economy*. Tubingen: J.C.B. Mohr.

Sen, A. (1995). *Inequality Reexamined*. Cambridge, MA: Harvard University Press.

Sen, A. (1999a). *Commodities and Capabilities*. Oxford: Oxford University Press.

Sen, A. (1999b). *Development as Freedom*. Oxford: Oxford University Press.

Sen, A. (2002). *Rationality and Freedom*. Cambridge, MA: Belknap Press.

Sen, A. (2009). *The Idea of Justice*. Cambridge, MA: Harvard University Press.

Sher, I. (2018). Evaluating Allocations of Freedom. *The Economic Journal 128*(612), F65–F94.

Shiffrin, S. V. (2000). Paternalism, Unconscionability Doctrine, and Accommodation. *Philosophy & Public Affairs 29*(3), 205–250.

Singer, P. (1973). Altruism and commerce: a defense of Titmuss against Arrow. *Philosophy & public affairs 2*(3), 312–320.

Singer, P. (1977). Freedom and Utilities in the Distribution of Health Care. In G. Dworkin, G. Bermant, and P. G. Brown (Eds.), *Markets and morals*. Washington: Hemisphere Publishing Corporation.

Steiner, P. (2015). The Organizational Gift and Sociological Approaches to Exchange. In P. Aspers and N. Dodd (Eds.), *Re-Imagining Economic Sociology*, pp. 275–298. Oxford: Oxford University Press.

Sugden, R. (1985). Liberty, preference, and choice. *Economics & philosophy 1*(2), 213–229.

Sugden, R. (2003). Opportunity as a space for individuality: its value and the impossibility of measuring it. *Ethics 113*(4), 783–809.

Sugden, R. (2006a). Taking Unconsidered Preferences Seriously. *Royal Institute of Philosophy Supplements 59*, 209–232.

Sugden, R. (2006b). What We Desire, What We Have Reason to Desire, Whatever We Might Desire: Mill and Sen on the Value of Opportunity. *Utilitas 18*(1), 33–51.

Sugden, R. (2007). The Value of Opportunities over Time When Preferences Are Unstable. *Social Choice and Welfare 29*(4), 665–682.

Sugden, R. (2010). Opportunity as Mutual Advantage. *Economics & Philosophy 26*(1), 47–68.

Sugden, R. (2018a). *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford: Oxford University Press.

Sugden, R. (2018b). What Should Economists Do Now? In R. E. Wagner (Ed.), *James M. Buchanan: A Theorist of Political Economy and Social Philosophy*, pp. 13–37. Basingstoke: Palgrave-Macmillan.

Sugden, R. (2020). Normative economics without preferences. *International Review of Economics 68*, 5–19.

Sunstein, C. R. (2014a). Choosing Not to Choose. *Duke Law Journal 64*(1), 1–52.

Sunstein, C. R. (2014b). *Why Nudge?: The Politics of Libertarian Paternalism*. New Haven: Yale University Press.

Sunstein, C. R. (2015). *Choosing not to choose: Understanding the value of choice*. Oxford: Oxford University Press.

Sunstein, C. R. (2019). *On Freedom*. Princeton: Princeton University Press.

Tabarrok, A. (1998). The private provision of public goods via dominant assurance contracts. *Public Choice 96*(3-4), 345–362.

Taylor, C. (2006). What's Wrong with Negative Liberty. In D. Miller (Ed.), *The Liberty Reader*, pp. 141–162. Edinburgh: Routledge.

Thaler, R. H. (2015). *Misbehaving: The Making of Behavioral Economics*. New York: W. W. Norton.

Thaler, R. H. and H. M. Shefrin (1981). An Economic Theory of Self-Control. *Journal of political Economy 89*(2), 392–406.

Thaler, R. H. and C. R. Sunstein (2003). Libertarian Paternalism. *American economic review 93*(2), 175–179.

Thaler, R. H. and C. R. Sunstein (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven: Yale University Press.

Thoma, J. (2021a). In Defence of Revealed Preference Theory. *Economics & Philosophy 37*(2), 163–187.

Thoma, J. (2021b). On the possibility of an anti-paternalist behavioural welfare economics. *Journal of Economic Methodology 28*(4), 350–363.

Titmuss, R. (2018). *The Gift Relationship: From Human Blood to Social Policy*. Bristol: Policy Press.

van Hees, M. (2022). Freedom. In M. Zwolinski and B. Ferguson (Eds.), *The Routledge Companion to Libertarianism*. Abingdon: Routledge.

Wegner, D. M., D. J. Schneider, S. R. Carter, and T. L. White (1987). Paradoxical Effects of Thought Suppression. *Journal of Personality and Social Psychology 53*(1), 5–13.

White, M. D. (2019). *The Oxford Handbook of Ethics and Economics*. Oxford Handbooks. Oxford: Oxford University Press.

Whitman, G. (2006). Against the New Paternalism. *Policy analysis 563*, 1–16.

Wight, J. B. (2015). *Ethics in Economics: An Introduction to Moral Frameworks*. Redwood City: Stanford University Press.

Contents (detailed)

Summary

This thesis focuses on the evaluation of economic policies and institutions, considering their impact on individual freedom. A new body of literature on the measurement of freedom of choice has emerged in economics in the last few decades, providing a framework for assessing how free individuals are within economic institutions. It assumes that an increase in freedom is valuable, even if it does not necessarily lead to greater happiness or better preference satisfaction. However, this perspective faces the obvious objection that individuals may not value increased freedom. In specific contexts, having too many options can be counterproductive, especially when time and attention are limited, issues of self-control arise, or in situations of strategic interaction. Individuals can improve their situation by deliberately restricting their options, a concept referred to as a 'commitment'.

This thesis proposes a paradigm shift by suggesting that choosing not to choose can be seen as an exercise of individual freedom rather than a renunciation of it. Economists like Amartya Sen, James Buchanan, Thomas Schelling, and Robert Sugden have already formulated or discussed this perspective, which will be explored further in the thesis. To do so, the nature of freedom implied in this judgment needs to be clarified. This requires a multidisciplinary perspective that combines economics and political philosophy, where freedom has been the subject of intense philosophical scrutiny. The paradoxical choice not to choose introduces philosophical complexities: this choice reveals a conflict in individual preferences, which contradicts the traditional economic assumption that they are stable and consistent. The following question arises: if individuals themselves sometimes prefer not to choose, can an intervention limiting their choices be desirable from the point of view of freedom? The answer hinges on how one represents the interests of the economic agent who makes this commitment and on one's view about freedom.

The thesis explores the methodological and ethical implications of these questions, highlighting the value judgments that economists makes when adopting one representation over another or selecting specific criteria for evaluating choices. Situated within the burgeoning field of philosophy of economics, the thesis investigates the methodological foundations of normative economics and the intricate intersection between ethics and economics. From the perspective of economic ethics, the thesis analyses the arguments that can be brought into various public debates on economics about the role of markets and the need for public interventions, drawing from economic literature and models. From a methodological perspective, it considers the role of values in normative economics as an evaluative discipline, contributing to a deeper understanding of the conceptual foundations underpinning normative economics.

Keywords

Freedom ; Normative economics ; Choice ; Paternalism ; Commitment ; Strategic interactions ; libertarianism ; Robert Sugden ; Amartya Sen

Résumé

Cette thèse porte sur l'évaluation des politiques et institutions économiques, en considérant leur impact sur la liberté individuelle. Une littérature sur la mesure de la liberté a vu le jour ces dernières décennies, fournissant un cadre pour évaluer à quel point les individus sont libres au sein des institutions économiques. Elle présuppose qu'une augmentation de la liberté est une bonne chose, même si elle ne conduit pas nécessairement à un plus grand bonheur ou à une meilleure satisfaction des préférences. Toutefois, cette perspective se heurte à l'objection évidente selon laquelle les individus peuvent ne pas accorder de valeur à une liberté accrue. Dans certains contextes, le fait d'avoir trop d'options peut être contre-productif, en particulier lorsque le temps et l'attention sont limités, que des problèmes de contrôle de soi se posent, ou dans des contextes d'interaction stratégique. Les individus peuvent améliorer leur situation en restreignant de manière délibérée leurs options, ce que j'appelle un acte d'engagement (*commitment*).

Cette thèse propose un changement de paradigme en suggérant que le choix de ne pas choisir peut être conçu comme un exercice de la liberté plutôt que comme un renoncement à elle. Des économistes tels qu'Amartya Sen, James Buchanan, Thomas Schelling et Robert Sugden ont déjà formulé ou discuté ce point de vue, qui sera approfondi dans la thèse. Pour ce faire, la nature de la liberté impliquée dans ce jugement doit être clarifiée. Cela implique une perspective multidisciplinaire alliant économie et philosophie politique, qui a fait de la liberté l'objet d'un examen approfondi. Le choix paradoxal de ne pas choisir introduit des complexités philosophiques : ce choix révèle un conflit dans les préférences individuelles, ce qui contredit l'hypothèse économique traditionnelle selon laquelle celles-ci sont stables et cohérentes. La question suivante se pose alors: si les individus préfèrent parfois eux-mêmes ne pas choisir, une intervention limitant leurs choix peut-elle être souhaitable du point de vue de la liberté ? La réponse dépend de la manière dont on représente les intérêts de l'agent économique qui fait cet acte d'engagement et de la conception de la liberté que l'on retient.

La thèse explore les implications méthodologiques et éthiques de ces questions, en mettant en évidence les jugements de valeur que les économistes émettent lorsqu'ils adoptent une représentation plutôt qu'une autre des intérêts de l'agent économiques ou qu'ils sélectionnent des critères spécifiques pour évaluer ses choix. Située dans le domaine en plein essor de la philosophie de l'économie, la thèse examine les fondements méthodologiques de

l'économie normative et l'intersection complexe entre l'éthique et l'économie. Du point de vue de l'éthique économique, la thèse analyse les arguments pouvant être présentés dans divers débats publics sur l'économie concernant le rôle des marchés et la nécessité d'interventions publiques, en s'appuyant sur la littérature et les modèles économiques. Du point de vue de la méthodologie, elle examine le rôle des valeurs dans l'économie normative, contribuant ainsi à une meilleure compréhension des fondements conceptuels qui sous-tendent celle-ci.

Mots-clés

Liberté ; Économie normative ; Choix ; Paternalisme ; Engagement ; Interactions stratégiques ; Libertarianisme ; Robert Sugden ; Amartya Sen

Résumé long

Historiquement, l'économie normative — la branche normative de l'économie, qui porte sur l'évaluation des situations économiques et non leur description, comme l'économie positive — a pris la forme de ce que l'on appelle l'économie du bien-être. L'insistance sur le bien-être est le produit de l'influence durable de l'utilitarisme sur l'histoire de la pensée économique, bien que la définition de ce qui compte comme bien-être ait considérablement évolué. Comme l'indique le nom, l'économie du bien-être utilise un critère de bien-être pour comparer et classer les alternatives (c'est-à-dire les options mutuellement exclusives accessibles à un décideur). Le bien-être ne signifie ici rien d'autre que la satisfaction des préférences individuelles — l'input de base des évaluations de l'économie du bien-être est constitué par ces préférences. L'économie du bien-être traditionnelle suppose que les individus peuvent classer les alternatives en fonction de leurs préférences et que ce classement est cohérent et stable dans le temps. Si les préférences individuelles ont ces caractéristiques, elles peuvent être représentées par une fonction d'utilité. Une autre caractéristique des préférences est qu'elles sont "révélées" dans les choix individuels, au sens où le choix que fait un individu parmi un ensemble d'alternatives est censé refléter son classement des alternatives selon ses préférences. Ce cadre est bien défini mais laisse sans réponse une question normative cruciale : pourquoi la satisfaction des préférences devrait-elle compter autant en matière d'évaluation des institutions et politiques économiques ?

Comme l'expliquent McQuillin et Sugden (2012), le cadre formel de l'économie du bien-être a le mérite de permettre la réalisation d'évaluations des politiques publiques (à travers des jugements sur l'optimalité au sens de Pareto et l'analyse coûts-avantages) tout en laissant ouverte l'interprétation philosophique du critère, ce qui rend le cadre accueillant à l'égard de diverses théories morales. Sugden et McQuillin énumèrent trois interprétations différentes qui ont été données du critère, et qui pourraient toutes justifier son adoption :

- La perspective utilitariste des premiers économistes néoclassiques tels que Pigou reposait sur la possibilité de mesurer l'utilité comme une quantité cardinale et inter-personnellement comparable, où l'utilité était comprise comme se référant à une sorte d'expérience hédonique. Dans l'économie du bien-être moderne, la satisfaction des préférences

peut être lue comme un indice (ordinal) pour mesurer les niveaux de bonheur individuel : plus quelqu'un a d'utilité, plus il est heureux. Cette "interprétation du bonheur" rapprocherait ainsi la satisfaction maximale des préférences de tout le monde du plus grand bonheur du plus grand nombre de Bentham.

- Selon l'"interprétation du bien-être", les préférences sont des indicateurs de ce qu'un individu "juge être son bien-être, ou de ce qu'il essaie d'atteindre". Plus un individu se rapproche de son alternative la plus préférée, mieux il se porte, selon son propre jugement, ce qui est désirable si on identifie le bien-être à la satisfaction des préférences subjectives.

- Selon l'"interprétation de la liberté" (ou de la souveraineté du consommateur), d'après la conception des préférences révélées, les préférences peuvent être vues comme un moyen de résumer les choix que les individus feraient, dans des circonstances appropriées. Comme les économistes prennent ces classements comme donnés sans faire de jugements à leur sujet, les individus sont libres de choisir ce qu'ils veulent vraiment.

Le problème est que lorsque "les préférences ne peuvent pas être supposées être cohérentes, ces différentes positions normatives ont des implications divergentes"[17]. Si, comme le suggèrent les résultats des économistes comportementaux montrant à quel point les incohérences dans le comportement de choix sont systématiques et répandues, les préférences (telles que révélées par les choix) ne peuvent pas être supposées cohérentes, il devient alors difficile de déterminer ce qui devrait constituer la base de l'évaluation, respectivement, de ce qui rend un individu heureux, de ce qui favorise son bien-être (tel qu'il le conçoit), de ce qu'il choisirait parmi un certain ensemble d'alternatives.

En tout état de cause, les économistes se trouvent confrontés à un "problème de réconciliation" — le "problème de la réconciliation de l'économie normative et de l'économie comportementale" (McQuillin et Sugden 2012). Le consensus superficiel qui soutenait l'économie du bien-être traditionnelle n'existe plus, car beaucoup d'économistes prennent maintenant au

[17]Toutes les traductions sont ici les miennes.

sérieux les résultats de l'économie comportementale qui remettent en question ses fondements. Comme le notent McQuillin et Sugden, de "nombreuses questions substantielles en philosophie morale et politique" qui "n'avaient pas besoin d'être posées", en raison de ce consensus, émergent maintenant. De nouvelles propositions ambitieuses, telles que la promotion par Sugden d'un "critère de l'opportunité" et son programme d'une "économie normative sans préférences" (Sugden 2021), ont gagné du terrain dans le domaine de l'économie normative. Le critère de l'opportunité, proposé par Sugden pour remplacer celui de satisfaction des préférences, fait appel à la valeur de la liberté :

> J'utilise le terme "opportunité" dans le sens qui est courant en économie et en théorie du choix social : une opportunité pour un individu est quelque chose qu'il a le pouvoir de réaliser, s'il le souhaite. (Sugden 2010).

L'utilisation d'un critère d'opportunité présente l'avantage, selon Sugden, de ne pas nécessiter de faire des hypothèses sur les préférences pour réaliser des évaluations — les individus qui ont plus d'opportunités sont simplement plus libres de choisir ce qu'ils préfèrent, quelles que soient leurs préférences. Une motivation pour introduire un critère de liberté ou d'opportunité est donc qu'il permettrait de résoudre le problème de la réconciliation.

Ces débats et évolutions des dernières décennies seraient impossibles à comprendre, cependant, si nous ne mentionnions pas l'absence traditionnelle d'engagement des économistes en matière de valeurs. Elle explique pourquoi les économistes acceptent le critère de satisfaction des préférences : il permet aux économistes de laisser la tâche de faire des jugements de valeur aux individus eux-mêmes. Cette position est caractérisée par Haybron et Alexandrova comme inspirée par un "minimalisme normatif" :

> Le minimalisme normatif est un ensemble de principes implicites de l'économie du bien-être. Il prétend réduire au minimum, voire éviter, les engagements en matière de valeurs, notamment en orientant l'économie normative uniquement vers la satisfaction des préférences, et ainsi (...) en s'en remettant aux jugements de valeur des individus (Haybron et Alexandrova 2013).

Les minimalistes normatifs adopteraient donc naturellement une position antipaternaliste — selon laquelle il revient à chaque individu de juger de ce

qui compte pour lui. L'antipaternalisme est généralement motivé par des valeurs telles que l'autonomie individuelle ou l'égalité. Ici, le refus de faire des jugements de valeur inspiré par une valeur de neutralité se transforme en une attitude antipaternaliste de déférence à l'égard des jugements de valeur propres des individus, ce qui rend le critère de satisfaction des préférences attrayant pour les économistes. Cependant, le virage comportemental en économie a également ébranlé le consensus superficiel sur l'antipaternalisme, en plus de l'économie du bien-être elle-même : ainsi Thaler et Sunstein, qui proposent de s'en remettre aux "vraies préférences" des individus, se qualifient aussi de "paternalistes libertaires" parce qu'ils proposent de modifier l'environnement des individus pour influencer leurs choix tout en laissant leur liberté de choix intacte. Ce nouveau "paternalisme comportemental", qui se retrouve également en philosophie (Conly 2013), a suscité une réaction de rejet de certains économistes et philosophes.

Il ne serait cependant pas exact de dire que les questions philosophiques sur le critère normatif de l'économie du bien-être n'ont pas été posées avant l'avènement de l'économie comportementale. Les écrits d'Amartya Sen peuvent être crédités d'avoir donné naissance à un vaste mouvement de travail interdisciplinaire consacré à la définition de moyens alternatifs d'évaluer les situations économiques. En particulier, l'approche par les capabilités met l'accent sur ce que Sen appelle les "libertés réelles" des individus, ce qui suppose de savoir les mesurer. Dans le sillage des travaux de Sen, un article pionnier de Pattanaik et Xu (1990) a marqué le début d'une nouvelle littérature (appelée la "littérature sur la liberté de choix") consacrée à la définition de règles pour classer des ensembles d'options, appelés "ensembles d'opportunités" (parfois aussi appelés "menus"), afin de déterminer à quelle condition un tel ensemble donnerait plus de liberté de choix à la personne qui choisit qu'un autre. Cette littérature tient pour acquis que la liberté a une valeur en elle-même, et qu'il est en principe possible de toujours augmenter la liberté en donnant plus de choix significatifs.

Supposons, comme le suggèrent Pattanaik et Xu (1990) que notre économie de marché soit remplacée par un système de commandement dirigé par des bureaucrates bienveillants et omniscients. Si les bureaucrates contraignent les individus à consommer le panier de biens qu'ils préfèrent le plus aux prix d'équilibre, les préférences des individus sont tout autant satisfaites que sous un système de marché, et donc les deux ensembles d'opportunités associés à chaque système sont également bons d'après le traditionnel critère de satisfaction des préférences de l'économie du bien-être. Et pourtant, dans ce

régime bureaucratique la liberté semble absente. Si les individus préféraient vivre dans une économie de marché plutôt que dans ce système, cela montrerait que la liberté leur importe en tant que telle, et non simplement parce qu'elle leur permet de mieux satisfaire leurs préférences. Ils se soucient aussi d'avoir des options sous-optimales — des options qu'ils ne choisiraient jamais — en plus de celles qu'ils préfèrent. Je dirai que, dans ce cas, les individus valorisent la liberté indépendamment (de sa capacité à satisfaire les préférences), ou qu'ils attribuent une valeur indépendante à la liberté. Deux types de justifications ont été avancées pour donner une valeur indépendante à la liberté : parce que la liberté permet aux individus d'exprimer leurs choix autonomes (justification kantienne), ou parce qu'elle leur permet d'exercer leur faculté de choix de façon à développer leur individualité et promouvoir leur bien-être global (justification millienne).

Que l'on donne ou non à la liberté une valeur indépendante, il existe des contextes où il n'y a aucune justification pour valoriser le fait d'avoir plus d'options — comme lorsque les individus ont une attention limitée, des problèmes de contrôle de soi, ou sont confrontés à des externalités et problèmes d'interaction stratégique. Dans ces contextes, les choix n'expriment pas la volonté autonome de celui qui choisit ou ne permettent pas à quelqu'un d'exercer ses facultés de choix pour développer son individualité. Par conséquent, dans ces contextes, le choix de ne pas choisir (de restreindre ses propres choix) n'est pas incompatible avec l'attribution à la liberté d'une valeur indépendante. Une préférence pour l'engagement, définie comme une préférence pour limiter l'extension de ses choix, peut avoir une certaine pertinence même pour quelqu'un qui se soucie du bien-être global ou de l'autonomie (et qui valoriserait donc la liberté de manière indépendante dans des contextes normaux). Mais la littérature sur la liberté de choix n'en dit pas beaucoup à ce sujet, car elle ne considère généralement pas les choix de choix. Plus qu'une lacune dans la littérature, il s'agit d'une frontière: que pourrait-on dire de la liberté dans ces cas ?

La thèse suit les traces d'économistes de renom qui, comme Sen, Schelling, Sugden, Buchanan, ont proposé de donner un sens, du point de vue d'une évaluation de la liberté, aux actes d'engagements par lesquels les individus restreignent leurs options. À ma connaissance, Sen et Sugden sont les seuls économistes à avoir décrit des critères normatifs précis qui reconnaissent et tiennent compte des engagements. Les notions d'"engagement stratégique" ou de "préférences pour l'engagement" sont bien connus des économistes, largement utilisées et appliquées. Cependant, peu d'efforts ont été déployés

pour interroger leur signification normative du point de vue de la liberté. Une préférence pour ne pas choisir a un caractère très paradoxal, ce qui en fait un défi pour examiner leur valeur d'un point de vue de la liberté. Cependant, elle est essentielle pour comprendre des débats importants ayant de larges implications sur la manière dont est structurée une économie.

Cette thèse relève de la philosophie de l'économie. Les philosophes et les économistes travaillent parfois sur les mêmes sujets, et parfois avec les mêmes méthodes. La philosophie de l'économie n'est pas seulement une extension du domaine de la philosophie des sciences (qui consiste à analyser les méthodes utilisées par les différentes sciences) à l'économie, car elle couvre également de nombreux domaines où l'économie et d'autres domaines de la philosophie, tels que la philosophie politique et éthique, se croisent. En tant que sous-discipline distincte de la méthodologie économique, elle est relativement nouvelle. Comme l'explique Hands (2018) :

> Depuis le milieu des années 1980, il y a eu une renaissance de l'interaction entre l'économie et la philosophie. L'approche traditionnelle de la méthodologie économique continue de produire des recherches viables, mais l'économie et la philosophie interagissent également de nombreuses autres manières nouvelles et importantes. (...) La dynamique intellectuelle est désormais celle d'une échange bilatéral (Hands 2008).

La thèse contribue à cet échange bilatéral de deux manières différentes. Du point de vue de l'éthique économique, la thèse analyse les arguments pouvant être présentés dans divers débats publics sur l'économie concernant le rôle des marchés et la nécessité d'interventions publiques, en s'appuyant sur la littérature et les modèles économiques. D'un point de vue méthodologique, elle examine le rôle des valeurs dans l'économie normative en tant que discipline évaluative

Présentation des chapitres :

Chapitre 1. Un problème auquel l'économie normative doit faire face est la nécessité de porter des jugements de valeur lors de la conception de modèles normatifs. Les modèles normatifs peuvent être définis comme "la classe de modèles formels visant à fournir des indications normatives" (Beck

et Jahn 2021). L'utilisation de jugements de valeur dans la conception et l'application de ces modèles a été peu explorée par la philosophie des sciences. En ce qui concerne l'économie, une exception intéressante est Sugden (2003), qui souligne qu'il est impossible de définir un modèle dit de "quantité pure" des opportunités individuelles, qui serait totalement neutre en ce qui concerne l'identification et l'évaluation des opportunités. Un modèle de quantité pure est censé refléter les opportunités disponibles dans le monde telles qu'elles sont et éviter tout jugement de valeur. Sugden suggère que c'est une tâche impossible. Je montre, dans un cas concret, pourquoi cela est le cas : selon ce que j'appelle "l'argument simple pour les marchés" (défendu en particulier par Arrow), l'ouverture de nouveaux marchés offre aux gens plus d'opportunités, quelles que soient leurs préférences. Avec un nouveau marché, ils peuvent désormais acheter et vendre là où ils ne pouvaient pas commercer auparavant, ou du moins pas contre de l'argent. Ce jugement ne devrait poser aucun problème du point de vue d'un modèle de quantité pure d'opportunités. Pourtant, si l'on revient sur les débats sur la marchandisation des transfusions sanguines qui ont opposé Titmuss et Arrow, on voit qu'ils tiennent des représentations incompatibles des opportunités disponibles aux individus après l'ouverture d'un marché de la transfusion sanguine. Je montre que ces deux représentations des opportunités sont liées à deux types différents de préférences distinctes (altruistes ou égoïstes). Selon que nous valorisons un type de préférences ou l'autre, la manière dont nous représentons les opportunités sera différente, ce qui signifie qu'il n'est généralement pas possible d'identifier les opportunités indépendamment des préférences que nous attribuons aux individus.

Chapitre 2. L'argument concluant à la nécessité pour l'Etat de fournir les biens collectifs est composé de trois prémisses distinctes : (1) les individus sont confrontés à un problème de bien collectif ; (2) il n'y a pas d'autre moyen de le résoudre que de recourir à une forme de coercition ; (3) l'amélioration de la situation de chacun qui résulte d'une intervention coercitive est suffisante pour compenser la perte de contrôle qu'elle implique. Les économistes n'ont généralement pas grand-chose à dire sur la troisième prémisse, même s'il s'agit d'une partie cruciale de l'argument. Les libertariens, qui s'opposent à la contrainte étatique autant que possible, peuvent refuser soit la prémisse (2) soit la prémisse (3) face à un problème de bien collectif. Certains d'entre eux ont fait valoir que la prémisse (2) est fausse car des "contrats d'assurance" peuvent être conçus et mis en œuvre volontairement pour coordonner la con-

tribution des individus au bien public et le produire. Je propose de considérer ce que j'appelle les "situations d'urgence", où ces contrats d'assurance ne sont pas réalisables en raison de la gravité de ces situations (par exemple, le début d'une épidémie mortelle, ou un réchauffement climatique rapide). Sur quels fondements les libertariens pourraient-ils alors refuser la prémisse (3) dans de telles situations ? Je défends l'idée, en suivant Sen, que les libertariens ont tort d'assimiler la liberté et le contrôle. Si la liberté peut exister là où il n'y a pas de contrôle, une intervention coercitive peut ne pas être perçue comme privant les individus de leur liberté. Si une intervention coercitive n'impose rien de plus aux individus que ce qu'ils se seraient imposé s'ils avaient accepté un "contrat d'assurance" pour fournir le bien collectif (et à condition qu'ils l'auraient accepté s'ils l'avaient pu), leur "liberté indirecte", comme l'on peut appeler après Sen, est préservée. Les libertariens qui reconnaissent l'attrait de cette notion de liberté indirecte peuvent accepter la prémisse (3) et l'idée que certaines interventions coercitives de l'État préservent la liberté.

Chapitre 3. Pour expliquer les phénomènes, les économistes utilisent une représentation du comportement de l'agent économique ainsi que de ses objectifs. Supposons que les individus soient vraiment tels que la représentation traditionnelle suppose qu'ils sont, ce qui signifierait que l'économie positive capture réellement les caractéristiques essentielles du comportement économique. Alors, comme l'a souligné Hausman (2021), en particulier, cela justifierait l'anti-paternalisme de l'économie du bien-être, car le paternalisme serait tout simplement sous-optimal. En effet, si (1) les individus choisissent ce qu'ils préfèrent, et (2) s'ils préfèrent ce qu'ils estiment être le meilleur pour eux, alors aucune intervention paternaliste, qui interférerait avec leurs choix, ne pourrait améliorer leur sort. Mais si au contraire les individus ne choisissent pas toujours, ou pas souvent, ce qui est le meilleur pour eux (comme le suggèrent certains économistes comportementaux), l'argument anti-paternaliste est réfuté. Dans le chapitre 3, je montre comment cette objection, avancée à la lumière des conclusions de l'économie comportementale, aurait pu être formulée bien plus tôt dans l'histoire de l'économie normative, sans renoncer aux prémisses (1) et (2), car il est bien connu, depuis que Thomas Schelling en a parlé, que, dans le contexte des interactions stratégiques, les individus ne choisissent pas, en un sens, ce qui est le meilleur pour eux. Cela a des implications importantes en matière d'intervention et de définition du champ d'action du paternalisme qui, elles, n'ont pas été reconnues : la possibilité de ce que j'appelle le "paternalisme stratégique".

Chapitre 4. L'anti-paternalisme en économie suppose de respecter les préférences des individus, quelles qu'elles soient. Cela implique qu'il faut une coïncidence entre ce qui importe pour les individus et le critère utilisé pour évaluer leur situation. Cependant, comme je le montre dans le chapitre 4, ce n'est pas le cas si nous adoptons un critère de liberté ou d'opportunité (comme le propose Sugden 2018a) pour évaluer les situations économiques et si nous admettons en même temps la possibilité que les individus aient une préférence pour l'engagement. Le chapitre met en lumière ce "trilemme", c'est-à-dire une incompatibilité générale entre l'anti-paternalisme, l'utilisation d'un critère de liberté et la possibilité d'une préférence pour l'engagement. Nous pouvons même aller plus loin et montrer qu'une position anti-paternaliste implique d'attribuer une préférence minimale pour la liberté aux individus. Il est donc impossible d'adopter une position anti-paternaliste "normative" sans devoir faire certaines hypothèses "positives" sur leur comportement. Pour le prouver, je considère un modèle le plus simple possible des interactions entre un citoyen et un économiste fournissant des recommandations aux autorités publiques. Le modèle montre qu'une position anti-paternaliste n'est cohérente que si les individus sont prêts à être traités comme tels, ce qui signifie qu'ils valorisent la liberté, dans un sens minimal.

Chapitre 5. Les dispositifs d'engagement (*commitment devices*) permettent aux individus de restreindre leurs choix futurs. Ce que l'on pourrait appeler des "dispositifs d'engagement forts" (DEF) le font en rendant matériellement coûteux le choix de certaines options, tandis que les "dispositifs d'engagement doux" (DED) le font en rendant psychologiquement coûteux ce choix. L'existence de marchés pour les DEF assure que les personnes ayant des problèmes de contrôle de soi peuvent en acheter pour s'empêcher de faire quelque chose demain qu'elles considèrent comme mauvais aujourd'hui. Les arguments que l'on peut avancer en faveur ou contre la possibilité d'acheter des DEF dépendent d'une certaine représentation des intérêts des individus enclins à des problèmes de contrôle de soi. Les économistes comportementaux ont proposé des modèles d'individus composés de multiples *moi* (*selves*) avec des préférences conflictuelles, qui ont été utilisés pour préconiser des interventions paternalistes. Sugden (2018a) a critiqué cette vision comme se basant sur l'idée non fondée que les individus devraient normalement avoir des préférences cohérentes, et que leurs

choix seraient cohérents s'ils n'avaient pas de problèmes de maîtrise de soi, une affirmation que la psychologie ne peut pas étayer empiriquement. Pour Sugden, si nous abandonnons cette idée et admettons plutôt que les individus sont des agents "responsables" qui considèrent leurs choix passés et futurs comme les leurs, il n'y aurait rien d'anormal à contrecarrer ses propres plans — de telles incohérences seraient simplement des indications que les gens ont changé d'avis. Cependant, le simple fait que les individus choisissent d'utiliser des DEF suggère qu'ils se voient eux-mêmes comme ayant plusieurs *moi* puisqu'ils anticipent qu'ils auront des préférences en conflit avec celles qu'ils ont maintenant. Ces deux représentations de l'agent semblent donc être empiriquement non fondées et insuffisantes pour justifier des conclusions normatives. Il y a place pour un point de vue différent qui ne ferait pas de telles hypothèses sur la psychologie de l'agent. Ainsi, du point de vue du principe de la liberté (*liberty principle*) de Mill, d'après lequel les individus doivent être laissés libre de faire ce qui ne nuit pas à autrui, ce qui compte est seulement si ce principe est respecté ou non, quelle que soit notre représentation des intérêts des agents. Or on peut montrer que ce principe ne peut pas justifier de forcer les individus à tenir leurs engagements contraignants envers eux-mêmes. Cette conclusion justifierait donc de réguler un marché des DEF pour limiter la prise d'engagements contraignants envers soi-même.

Cette thèse porte sur l'évaluation des politiques et institutions économiques en considérant leur impact sur la liberté individuelle. Des économistes tels que Sen et Hayek ont mis en avant l'importance de la liberté. Depuis, la question de la mesure de la liberté de choix a suscité un grand intérêt. Cependant, trop de choix peut être jugé contre-productif, justifiant la nécessité de limiter ses options. La thèse avance que choisir de ne pas choisir peut être un acte de liberté, plutôt qu'un rejet de celle-ci. Ce choix paradoxal soulève des questions philosophiques, la façon dont on y répond dépend de la conception de la liberté et de la représentation des intérêts de l'agent économique que l'on adopte. La thèse en examine les implications pour l'économie. Elle relève du domaine de la philosophie de l'économie, explorant l'intersection entre l'éthique, la philosophie des sciences et l'économie. Elle examine le rôle des valeurs en économie normative, contribuant ainsi à une meilleure compréhension de ses fondements conceptuels.

This thesis focuses on the evaluation of economic policies and institutions, considering their impact on individual freedom. Economists like Sen and Hayek have emphasized the importance of freedom. Since then, the question of measuring freedom of choice has sparked significant interest. However, too much choice can be deemed counterproductive, justifying the need to limit one's options. The thesis argues that choosing not to choose can be an exercise of freedom rather than a rejection of it. This paradoxical choice raises philosophical questions, the answers to which depend on the conception of freedom and the representation of the economic agent's interests we adopt. The thesis examines its implications for economics. It belongs to the field of philosophy of economics, exploring the intersection between ethics, philosophy of science, and economics. It delves into the role of values in normative economics, contributing to a better understanding of its conceptual foundations.