The Case Against Self-Constraint

Kevin Leportier

The Case Against Self-Constraint

Kevin Leportier*
University Paris 1 Panthéon-Sorbonne
Centre d'économie de la Sorbonne

May 7, 2024

Abstract

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 paper, 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.

Keywords: Freedom, Self-Control, Commitment, Paternalism, John Stuart Mill

*e-mail address : kevin.leportier@gmail.com

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 2018). 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 paper 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 paper 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.

The goal of this paper 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.

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

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

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 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

[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.

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'

[3]See Thaler and Sunstein (2008).

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

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 of paternalism given in paper 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

[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.

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 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 paper 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

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

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.

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.

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.

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.

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 internalization 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.

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.

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

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 provides 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

[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)

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 considerations 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 2018, 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 2018, 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

[10]See Fumagalli (2023, 11).

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 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

[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.

self that is constrained. (Sugden 2018, 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 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.

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

[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.

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,

> 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.

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

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 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.)

[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.

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 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.

References

Ainslie, G. (1992). *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person*. Cambridge: Cambridge University Press.

Andreou, C. (2022). *Commitment and Resoluteness in Rational Choice*. Cambridge: Cambridge University Press.

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.

Bryan, G., D. Karlan, and S. Nelson (2010). Commitment Devices. *Annual Review of Economics 2*(1), 671–698.

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.

Descartes, R. (2015). *The Passions of the Soul and Other Late Philosophical Writings*. Oxford: Oxford University Press.

Dworkin, G. (1972). Paternalism. *the Monist 56*(1), 64–84.

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.

Fumagalli, R. (2023). Preferences Versus Opportunities: On the Conceptual Foundations of Normative Welfare Economics. *Economics & Philosophy*, 1–25.

Gul, F. and W. Pesendorfer (2001). Temptation and Self-Control. *Econometrica 69*(6), 1403–1435.

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.

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.

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.

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.

Mill, J. S. (2006). *On Liberty and The Subjection of Women*. London: Penguin Classics.

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'Donoghue, T. and M. Rabin (2006). Optimal sin taxes. *Journal of Public Economics 90*(10), 1825–1849.

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.

Saint-Paul, G. (2011). *The Tyranny of Utility: Behavioral Social Science and the Rise of Paternalism*. Princeton: Princeton University Press.

Schelling, T. C. (1978). Egonomics, or the Art of Self-Management. *The American Economic Review 68*(2), 290–294.

Schelling, T. C. (1984). *Choice and Consequence*. Cambridge, MA: Harvard University Press.

Strotz, R. H. (1956). Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies 23*(3), 165–180.

Sugden, R. (2003). Opportunity as a space for individuality: its value and the impossibility of measuring it. *Ethics 113*(4), 783–809.

Sugden, R. (2018). *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford: Oxford University Press.

Sunstein, C. R. (2014). *Why Nudge?: The Politics of Libertarian Paternalism*. New Haven: Yale University Press.

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 (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven: Yale University Press.

Whitman, G. (2006). Against the New Paternalism. *Policy analysis 563*, 1–16.