Economics and Philosophy, **31** (2015) 275–295 © Cambridge University Press
doi:10.1017/S0266267115000139 First published online 5 May 2015
journals.cambridge.org/eap

OPPORTUNITY AND PREFERENCE LEARNING

CHRISTIAN SCHUBERT*

Abstract: Robert Sugden has suggested a normative standard of freedom as 'opportunity' that is supposed to help realign normative economics – with its traditional rational choice orientation – with behavioural economics. While allowing preferences to be incoherent, he wants to maintain the anti-paternalist stance of orthodox welfare economics. His standard, though, presupposes that people respond to uncertainty about their own future preferences by dismissing any kind of self-constraint. We argue that the approach lacks psychological substance: Sugden's normative benchmark – the 'responsible person' – can hardly serve as a convincing role model in a contractarian setting. An alternative concept is introduced, and some implications are briefly discussed.

Keywords: Opportunity criterion, preference change, reconciliation problem

1. INTRODUCTION

In a series of papers, Robert Sugden has elaborated upon the case for a criterion of freedom as 'opportunity' that is supposed to cope with the problem of how to realign normative economics – with its traditional grounding in orthodox rational choice methodology – with behavioural economics (e.g. Sugden 2004a, 2006b, 2007, 2008a, 2010; McQuillin and Sugden 2012).[1] Among the various attempts that have been made in the literature to come to terms with this 'reconciliation problem'

* Institute of Economics, University of Kassel, Nora-Platiel-Strasse 4, 34109 Kassel, Germany. Email: dr.c.schubert@gmail.com. URL: http://www.uni-kassel.de/fb07/institute/ivwl/faculty-chairs/schubert/home.html

[1] See also Sugden (1998a, 1998b, 2003, 2004b, 2006a, 2008b, 2009, 2013). In general, there is a growing emphasis on opportunity (rather than preference satisfaction) as a 'currency of advantage' among normative economists, see e.g. Sen (1992), Arrow (1995) and Roemer (1998).

(McQuillin and Sugden 2012) and to explore the normative implications of behavioural economics – such as 'libertarian paternalism' (e.g. Camerer *et al*. 2003; Thaler and Sunstein 2008), 'behavioural welfare economics' (e.g. Bernheim and Rangel 2009), and 'happiness politics' (e.g. Schubert 2012) –, his stands out in two respects.[2]

First, Sugden wishes to refute the claim that acceptance of behavioural economics insights makes anti-paternalism unattractive or even 'incoherent' (Sunstein and Thaler 2003: 1164–1165, 1182). His point of departure is the failure of the standard welfarist approach to normative economics to work properly, once it is clear that individuals often do not reveal coherent (i.e. consistent, stable and context-independent) preference orderings through their observable choices. The opportunity criterion that he proposes allows the possibility of incoherent preferences, while retaining the substance of liberal welfare economics – i.e. the principle of consumer sovereignty. Sugden's underlying normative intuition is that 'it is good that each person is free to get what she wants', provided she does not harm others (Sugden 2004a: 1016). Being free to choose from a wide range of options, whether or not one's choices reveal coherent preferences, is taken to be valuable (McQuillin and Sugden 2012: 563). The microfoundation is provided by the concept of a 'responsible' (rather than 'rational') person that fully endorses the decisions made by all her – past and future – selves (Sugden, 2004a: 1018). Such a person will value opportunity because, anticipating future periods, she wishes to be free to get what she then happens to desire (Sugden 2010: 55).

Second, and apart from the substantive suggestion to adopt the criterion of opportunity, Sugden proposes to redirect normative economics more generally in a contractarian direction: Rather than addressing some supposedly benevolent social planner, normative economists should try to convince the citizens directly. The concept of opportunity is, thus, taken to be offered to them as part of a 'fair agreement', to wit, a currency of advantage in which to assess specific institutions and policies (see in particular Sugden 1989, Sugden 2013).

As we will argue, despite all its merits, Sugden's concept of opportunity suffers from being based upon an incomplete model of the attitude real-world individuals are likely to assume toward the fact that their own preferences will change in future periods. Instead of allowing the possibility that individuals wish to selectively indulge their own future preferences (in order to manage the way they develop), he narrows the range of attitudes down to just one: Unconditional endorsement. As a consequence, Sugden will most likely fail to convince his addressees

[2] The capability approach (e.g. Sen 1988) might also be interpreted as an attempt to tackle the reconciliation problem, but since it is rarely explicitly justified on this ground, we abstract from it here.

that his concept offers a plausible basis for a 'fair agreement'. Crucially, he downplays the practical relevance of personal precommitment: 'Most of us, most of the time, recognise that the best way to cope with ... uncertainty [about one's future preferences] is to avoid making unnecessary commitments about future consumption' (Sugden 2008a: 243). Only by identifying with – and privileging as authoritative – the 'acting' or 'impulsive self' (in his role as consumer, say), rather than the 'planning self' might one be able to appreciate the value of the market in providing for opportunity (Sugden 2008a: 243).

Our argument proceeds as follows. Section 2 presents, in a nutshell, Sugden's concept of opportunity, as he has developed it to date. Section 3 discusses what are, in our view, its main shortcomings. Section 4 then introduces an alternative notion of 'opportunity to learn' and discusses some implications regarding, for instance, the use of some of the tools advocated by 'libertarian paternalists'. Section 5 concludes.

2. SUGDEN'S CONCEPT OF OPPORTUNITY

Sugden aims at developing a 'standard of value' that does not require individuals to hold stable and coherent (or 'considered') preferences. It is supposed to allow for preferences to be formed in a not necessarily 'rational' way in the very process of interaction on markets for private goods (Sugden 2008a: 230).[3] Hence, such a standard would be applicable in a world such as the one described by behavioural economics. At the same time, Sugden wishes to preserve, as much as possible, the (classical) liberal emphasis on personal freedom – epitomized in the principle of consumers' sovereignty.[4] Specifically, he claims that although individual preferences that are being satisfied may themselves be unconsidered, and, therefore, not suitable as a basis for an account of subjective value, people's 'valuing the *opportunity to satisfy them* is considered' (Sugden 2006b: 217, italics partly omitted).[5] In other words, while the opportunity to satisfy preferences may be *used* in an unconsidered way (e.g. leading to dynamic inconsistencies), the valuation of this opportunity itself can be assumed to be perfectly coherent and stable. Thus, Sugden suggests substituting an opportunity criterion for the traditional preference satisfaction standard: Opportunities, rather than

[3] He applies his criterion to the provision of public goods in Sugden (2009).

[4] The defence of the principle of Consumer Sovereignty by Sugden is somewhat puzzling, given that it is firmly embedded in a welfarist framework of preference-satisfaction that Sugden actually rejects. We submit that what he really means to defend is the principle of *freedom of choice*. Both principles can have conflicting implications (Rothenberg 1962).

[5] Generally, in social choice theory, opportunity for a given individual is defined as 'something that he has the power to bring about, if he so chooses', a definition Sugden endorses (Sugden 2010: 49).

outcomes or consumption bundles, are then the ultimate carriers of value. Given his contractarian framework, he has to show (i) that his claim is 'credible as moral psychology' and (ii) that the corresponding standard of value can be formalized in a coherent way (Sugden 2006b: 218). In elaborating upon his approach, Sugden has so far focused on question (ii), answering question (i), affirmatively, basically by introspection (see e.g. Sugden 2007: 671; Sugden 2008a: 243, 247).

Sugden demonstrates that under certain ideal conditions, competitive markets satisfy the opportunity criterion, i.e. they provide individuals with maximum opportunities (Sugden 2004a).[6] By providing maximum chances for mutually beneficial transactions (including intertemporal transactions between an agent's own selves), markets give each person, 'rational or irrational, what she wants and is willing to pay for, when she wants it and is willing to pay for it' (McQuillin and Sugden 2012: 630–631). In a competitive market setting with free trade, each person's opportunities are a function of other people's desires – which implies that the opportunities of all persons are potentially connected in a network of positive-sum interactions (Sugden 2010: 55–62). The key is that the market gives individuals incentives to try to predict (and try to satisfy) what other individuals will prefer in future periods (Sugden 2008a: 244). Sugden even extends this argument to the case of preferences *created* by market participants through, e.g. advertisement (Sugden 2008a: 244–247). Hence, he specifies his concept as 'opportunity as mutual advantage' (Sugden 2010).[7]

His methodological background assumptions are most lucidly described in Sugden (1998b). There, he explicitly adopts a position that he attributes to Hume and Gauthier (and indirectly to Pareto), according to which an individual's actual preferences are not the proper subject of rational assessment in light of reason. Preferences should rather be understood as 'passions'. The underlying subjective beliefs may well be flawed, but the preferences themselves can never be 'irrational'. This implies the rejection of any kind of substantive theory of well-being (such as happiness or capability). In particular, a person's preference for A over B should not be interpreted as some kind of overall all-things-considered

[6] In more abstract terms, markets can 'integrate fragments of preference that are revealed in different situations and at different times', just as they can integrate fragments of dispersed knowledge (Sugden 2008a: 229, 239). As he shows in (Sugden 2004a), in a simple model of an exchange economy, the opportunity criterion is satisfied if the law of one price holds and if all markets clear. The model presented there is extended in McQuillin and Sugden (2012) and Sugden (2014).

[7] This reflects Hayek's insight that 'it is because every individual knows so little ... that we trust the independent and competitive efforts of many to induce the emergence of what we shall want when we see it' (Hayek 1989: 55–56).

reason for choosing A over B; it does not have to be justified to anyone else – particularly not to an ethical observer.

From this point of view, it is easy to dismiss the welfarist approach to well-being as incoherent: Even if preferences were to turn out as stable and consistent as postulated in the textbook models, there is no ground for assuming that a person's good always equals the satisfaction of her preferences (Sugden 1998b: 41). The ethical observer's judgements about what is good for a person are based on reasons, while the person's preferences are not. While dismissing welfarism, though, Sugden wishes to keep the welfarists' neutrality with respect to accounts of well-being. He endorses Harsanyi's principle of preference autonomy to work with each individual's own conception of well-being (Sugden 1998b: 46). In order to do so, he adopts a contractarian approach.

2.1 The 'responsible agent'

While eschewing any substantive theory of well-being, Sugden nonetheless needs a conception of individual interest (Sugden 1998b: 60): What does each agent seek to achieve from social cooperation? At this point, he introduces his critical premise, viz. that each individual seeks to maximize the extent of her opportunities. This is supposed to offer a 'reasonably realistic' model of individuals' judgements about opportunity (Sugden 1998b: 55). Given this assumption about people's 'pervasive' preference for increases in opportunity, he then classifies this preference as a passion, which therefore does not require any justification in terms of a particular theory of well-being – it does not make sense to ask *why* someone benefits from increases in her opportunity set (Sugden 1998b: 53). Endowed with this specification of individual interest, he can hypothesize about what kind of institutional arrangements each individual, taken separately, can agree to.

In order to back the underlying empirical claim that there is indeed, among real-world individuals, a widely held 'passion' for maximizing one's own opportunity sets, Sugden suggests a specific conception of personhood or identity that also serves as a normative role model: For him, identity consists 'not in a coherent system of preferences or in a rational plan of life, but in an *attitude towards one's past, present and future actions* which acknowledges those actions as one's own and accepts responsibility for their consequences' (Sugden 2010: 54–55, italics added). His normative ideal is the *responsible*, rather than the rational individual. A responsible person, in Sugden's view, treats her past, present and future actions as her own, even when faced with regret or uncertainty (Sugden 2004a: 1018). Couched in the multiple-selves language, a responsible agent is a *continuing* agent, i.e., a 'composition of the series of time-slice agents' (Sugden 2007: 671, italics omitted). Individuals acting on incoherent and unstable preferences are, then, not treated as consisting of multiple selves with conflicting desires (of whom one ought to be privileged), but as continuing loci of responsibility. They endorse any desire their selves once had or will have – they even *identify* with their separate selves (Sugden 2007: 672, fn. 5).[8]

Sugden has formalized his intuition about what 'responsibility' entails by using the concept of 'nested opportunity sets' to model different multi-period decision problems for an agent (Sugden 2006b, 2007). Conceptualizing each person's opportunity as her set of 'allowable lifetime behaviours' (McQuillin and Sugden 2012: 628), he shows that, given the inclination to identify with one's own selves, more opportunity is unambiguously preferable to less (even if the extra opportunity turns out to generate unambiguous losses in terms of well-being), and tools to reduce opportunities in future periods have zero value.

2.2 Self-command versus self-constraint

Crucially, Sugden distinguishes between two kinds of self-commitment, namely, 'self-constraint' and 'self-command'. The former denotes the imposition of external restrictions on one's future choices, such as locking away spirits in a cupboard and sending the key back to oneself. The latter, by contrast, denotes a purely willpower-induced personal resolution to refrain from certain choices (such as heavy drinking) in the future.[9] As Sugden emphasizes, only what he refers to as self-*command* can be positively valued by agents endorsing his opportunity criterion.[10] Self-*constraint*, by contrast, will be negatively valued by a responsible agent, due to the 'objective' reduction in opportunities it implies. In theoretical terms, Sugden rejects the multiple-selves models typically used to show that under certain circumstances, 'external' self-constraints may be individually beneficial. For to show that, those models usually – and somewhat arbitrarily – privilege one (typically the 'long-term') self over others. An essential implication of Sugden's approach is that the tools suggested by 'libertarian paternalism' potentially impose normative costs – largely overlooked in the literature – by introducing external constraints that curtail the valued range of opportunities.[11]

[8] At first sight, there is a striking resemblance of this role model with the aristocratic attitude of 'never complain, never explain', usually attributed to Benjamin Disraeli. Note the normative nature of that dictum.

[9] The examples are Sugden's (personal communication, March 2013). In the following, we will use 'self-command' and 'self-constraint' in the sense suggested by Sugden, and 'self-commitment' as encompassing both.

[10] Personal communication (March 2013). See also Sugden (2004a: 1018; 2007).

[11] Specifically, Sugden wants to counter the impression that accepting the insights of behavioural economics makes anti-paternalism 'incoherent', see Sugden (2008a: 229),

2.3 How to measure opportunity

Suggesting opportunity as the standard of value prompts the question of how to measure it – in order to determine a 'fair' distribution, say. Sugden opposes both a purely quantitative approach – that measures opportunities and their value independently of preferences – and any approach that takes agents' preferences as given. Instead, the metric of opportunity should, in his view, be based on people's *potential* (counterfactual) preferences, i.e. those that an agent might adopt, given her 'objective circumstances', but assuming that her preferences are still unformed (Sugden 2003: 791; 2010: 50). Specifically, he eschews the approaches, common in the literature, to rely on 'reasonable' or 'typical' preferences – Sugden wants to allow the individuals to have eccentric preferences as well.[12] Given these strictures, though, it is impossible to measure a person's opportunity directly. In Sugden's view, opportunity rather becomes an 'open-ended' concept (Sugden 2010: 48). In order to be nevertheless able to use the concept in a meaningful way, he suggests to approximate the benefits one can derive from a given set of opportunities as a function of the resources or real income an agent possesses (Sugden 2010: 49). Specifically, the distribution of opportunities can only be analysed by measuring *entitlements* (Sugden 2010: 62–63). Then, the notion of opportunity requires neither the assumption of rationality nor committing to any particular philosophical conception of the good life. One can make judgements about the relative value of alternative opportunities without knowing anything about the substance (let alone the coherence) of one's own future preferences.

We will now examine what is wrong with this account of opportunity, and then suggest an alternative concept.

3. THREE SHORTCOMINGS OF THE OPPORTUNITY CONCEPT

In what follows, we will challenge Sugden's claim that the standard of value which he proposes represents a 'reasonably realistic' model of people's judgements about opportunity (Sugden 1998b: 55) and, hence, builds upon 'credible moral psychology' (Sugden 2006b: 218). Notice that we join Sugden in adopting a contractarian perspective. This implies two

McQuillin and Sugden (2012). Sunstein and Thaler (2003: 1167, fn 22), for example, advance the strong claim that 'the autonomy argument [meaning the 'belief that people are entitled to make their own choices even if they err'] is undermined by the fact … that sometimes preferences and choices are a function of given arrangements.' From observing that preferences do not 'predate social contexts' they conclude that there are no 'viable alternatives to paternalism' (Thaler and Sunstein 2008: 235; see also Sunstein and Thaler 2003: 1164–1165, 1182).

[12] Sugden defines preferences as 'dispositions to make specific choices … the product of a process of deliberation in which the agent decides what to choose' (Sugden 1998a: 323).

things: first, to address citizens 'as they are', i.e. refraining from imposing some specific substantive theory of well-being; second, to aim at a model of man that is psychologically plausible.

We argue, first, that Sugden's distinction between 'self-command' and 'self-constraint' is unconvincing. Second, his claim about 'most' individuals wishing to maximize their opportunity sets as these present themselves at every moment not only lacks empirical backing per se, but also neglects evidence from behavioural economics about agents' demand for self-constraint: agents may use self-constraint in order to avoid choice overload. Choice overload points toward the fact that opportunities are subject to intertemporal trade-offs. Third, the demand for self-constraint can also be motivated by an underlying desire to maintain subjective coherence among one's own preferences over time. While the first argument is of a conceptual nature, the second qualifies as a 'strong' one, for it is independent of any substantive account of well-being (in other words, it is firmly contractarian); the third one, however, qualifies as 'weak' in the sense that it does presuppose a substantive account of well-being.

3.1 Tools of self-commitment

The key empirical assumption backing Sugden's 'responsible agent' rests on the distinction between self-command and self-constraint. Responsible agents are said to accept the former, but reject the latter (see above). This raises the following problem, though: Even analytically, it is hard to separate these two tools of self-commitment. Let's look again at the examples brought up by Sugden himself (personal communication, March 2013): Consider someone who resolves to stay sober at an upcoming party. This exemplifies self-command. Alternatively, that person might wish to lock away the drinks in a cupboard and send back the keys to herself by second-class mail. That is supposed to qualify as self-constraint. However, the difference between both types (or tools) of self-commitment is only in the costs self-imposed by the agent concerned. Hence, it seems to be gradual only. Consider 'self-command': The party guest could, for example, strengthen her self-command by telling other guests about her resolution, thereby raising the 'moral costs' incurred by drinking alcohol. Before telling anyone about her resolution, these moral costs only included the psychic costs of shame and lowered self-esteem, say. 'Self-constraint' is equally fuzzy: By incurring some physical costs (using a screwdriver, say), she could opt for breaking open said cupboard. Depending on the circumstances, she may wish to vary the degree of self-commitment in the sense described here. Being gradual only, the distinction between self-command and self-constraint then hinges on the theorist's (ultimately arbitrary) decision on where to draw the line, either in terms of the level or the kind of costs imposed.

Apart from the blurry distinction between both types of self-commitment, it is also unclear why exactly 'imposing external restrictions on one's opportunity set' (i.e. self-constraint) should be incompatible with the opportunity criterion – or, equivalently, why a 'responsible' agent in Sugden's sense would never wish to consider her own planning self's reflective or 'second-order' preferences, alongside those of her acting or 'impulsive self'.[13] According to Sugden, the opportunity criterion requires agents to 'live with the consequences' of whatever preference their future selves end up having. But why insist that it's only the 'spontaneous' *first-order* preferences that count? Developing a preference for restricting one's own options seems to be entirely consistent with acting 'responsibly' in this sense, *as long as it is the agent herself* who voluntarily chooses to self-commit, perhaps as part of her overall attempt to manage the development of her preferences (see below, section 4), and as long as she accepts the possibility and the consequences of failed attempts of self-commitment. Then, we could speak of 'responsible self-commitment'. Put differently, privileging the impulsive over the planning self seems just as arbitrary as, for instance, privileging long-term over short-term preferences, a strategy that is common in multiple-selves models and rightly criticized by Sugden (e.g. Sugden 2006b; see Read 2006 on the intricacies of these kinds of implications of multiple-selves models).

3.2 The costs of choosing

What is it that might motivate agents to act responsibly in that extended sense, i.e. to listen to their planning self and ask for self-constraint? It is by now a well-established fact in behavioural economics that the ongoing increase of options and, hence, opportunities in Sugden's sense, in most domains of life (products, services, jobs, institutional arrangements) does not necessarily increase people's well-being. Rational individuals, of course, would only ever consider the one single next-best alternative to be factored in their personal cost-benefit calculus and, hence, unambiguously benefit from expanding opportunity sets. Real-world individuals, though, often lack the capacity to do so and tend to overestimate the combined opportunity costs of their choices. Coupled with loss aversion, they feel cognitively overwhelmed, ending up with lower levels of well-being than before (Schwartz 2004: ch. 6).

[13] The concept of second-order or meta-preferences (representing preferences over preferences) is due to Frankfurt (1971). Sugden's sceptical view on it is mainly motivated by the popular strategy to identify people's (allegedly superior) 'long-term' preferences with their meta-preferences, see e.g. Sugden (2006b).

Given our aim to develop an 'internal' – contractarian – critique, we of course cannot adduce any argument that is based on a substantive theory of well-being. Remember, we have to 'take people as they are'. There is evidence that supports our case even under these conditions of strict neutrality, though: Agents confronted with expanding opportunity sets tend to 'choose not to choose', i.e. to avoid decisions altogether (e.g. Iyengar and Lepper 2000; Shin and Ariely 2004). That, however, is tantamount to a reduction of opportunities. Thus, it seems that it does not make sense to simply postulate, as Sugden does, a pervasive preference for maximizing opportunities. Rather, the amount of opportunities real-world agents can handle seems to be limited. This has an important implication: There seems to be a trade-off between opportunity sets at different points in time: Maximizing one's opportunities in the short term may decrease the amount of opportunities one is able to enjoy in the long term – which makes it more plausible to see individuals as being engaged in dynamic optimization rather than static maximization when it comes to opportunities. Optimizing one's opportunities over time, though, requires some self-commitment, including self-constraint, in order to manage the amount of opportunities present at any given time. In other words, not only can real-world agents be *better off* by embracing some voluntary self-constraint (Schwartz 2004: 5), but some constraints of this kind can also be helpful in managing the currency of advantage we are concerned with here, viz. opportunity. Hence, this argument qualifies as a 'strong' one in our context: While an agent who only cares about the extent of her opportunities would accept (and 'live with the consequences of') losses in terms of well-being, she would not accept situations that are harmful in terms of opportunity losses for her future selves. For that would run counter to her postulated contractarian interest.

What all this shows is that that real-world agents will often be motivated to consider their planning selves' preferences: Any decision not to choose among given opportunities reflects those preferences. This puts into doubt the plausibility of Sugden's role model of the 'responsible' person who only ever endorses what his impulsive self wants.

3.3 Striving for subjective coherence

There is an additional factor that might motivate self-constraint among the addressees of Sugden's contractarian advice. This factor, though, presupposes, contra Sugden, that real-world agents do pursue *some* conception of well-being. Why is this nevertheless a valid argument in our context?

Sugden's opportunity concept is meant to convince a very broad class of individuals by appealing to the – allegedly 'pervasive' – preference for opportunities, rather than to any specific conception of well-being. In other words, its validity is supposed to be independent of any particular conception of well-being. Our following argument, though, presupposes 'most' real-world individuals to pursue *some* conception of well-being, be it wealth, happiness or whatever. It does not apply to those who care only about their opportunities. Thus, we propose it as a 'weak' argument only.[14]

The claim, then, is that to the extent that real-world individuals do care about their well-being, they strive for a minimum degree of subjectively perceived coherence between their personal beliefs and preferences, both at any given moment and over time.[15] In order to achieve this, they require tools of both self-command and self-constraint. As support for this claim, we can adduce evidence from behavioural economics about the human drive to reduce cognitive dissonance (e.g. Akerlof and Dickens 1982),[16] to preserve one's preferences over time (Sally 2000), to consolidate one's preferences once they have been constructed (Hoeffler and Ariely 1999), and to counter the effects of utility misprediction (Loewenstein and Angner 2003).

While in the literature, the demand for self-commitment devices is usually seen as being driven by the desire to maintain objective internal consistency, i.e. rationality in the neoclassical sense,[17] we argue that it should rather be understood as expressing the desire to maintain a minimum degree of coherence, *as perceived by the individual herself*. What the individual perceives as coherent may look perfectly incoherent ('eccentric') from an observer's viewpoint. Typically, willpower (i.e. self-command) has to be mustered to achieve a minimum degree of subjective coherence (Loewenstein 1999), but it is hard to see why agents should, in principle, refrain from supplementing willpower with some self-constraint.

[14] Recall Sugden's own twofold condition for valid contractarian advice, above. Within his own framework, both conditions boil down to just one: real-world individuals only care about their opportunity sets.

[15] This includes efforts at obstructing the possibly negative consequences of subjectively perceived *incoherence*.

[16] The psychological theory of cognitive dissonance describes the human disposition to adapt one's *beliefs* to one's changing preferences. Since Sugden assumes agents to be 'responsible' for their own preferences, he must postulate that they are also able to control their preferences, at least partly. For the agent, this may be preferable than adjusting one's beliefs when, e.g., there is a strong desire for certain beliefs (see e.g. Akerlof and Dickens 1982).

[17] See e.g. Schelling (1978, 1984) and O'Donoghue and Rabin (1999). Note that a desire to maintain internal consistency in the economics textbook sense can hardly be given motivational underpinnings. For instance, it is doubtful whether it makes individuals systematically better off than seemingly 'irrational' behaviour in strategic interactions (see e.g. Berg and Gigerenzer 2007). See, however, Yaari (1977).

As shown above, commitment by self-command can be gradually turned into what Sugden refers to as 'self-constraint', and possibly qualify as an even more useful tool to manage one's own preferences: By removing one's cigarettes, for instance, one may try to force one's own first-order preferences into a desired direction, i.e. to align them with one's second-order preferences. Assuming that preferences are partly shaped by what's on offer in the marketplace, the concept of second-order preference also allows us to show why the market cannot be expected to bring about, on its own, the kind of subjective coherence people in fact desire. Changes in economic circumstances may bring about situations where my own first-order preferences are 'improved' (in the sense of being more aligned with my own second-order preferences), but since this has all the characteristics of a positive externality, the market will likely underperform in this regard (George 2004, 2006).[18] Note that this argument does not presuppose any *specific* theory of the good life – the agent in question may be driven by any kind of evaluative commitments.

These considerations all lend substance to our claim that even within a contractarian perspective that takes people 'as they are', the opportunity criterion proposed by Sugden lacks psychological credibility, in that it is based on an implausible assumption about the attitude most people are likely to have toward the evolution of their own future preferences. Sugden's 'responsible agent' cannot, then, serve as a convincing normative role model to guide contractarian advice.

4. TOWARD A CONCEPT OF 'OPPORTUNITY TO LEARN'

In light of these objections, we suggest an alternative concept that may be understood as a dynamic variant of Sugden's: let's call it 'opportunity to learn'. In this section, we will describe the underlying normative intuition, along with the key empirical hypotheses that support it. Then, we will defend our concept against two possible objections: We show, first, that the account of preference learning it dwells upon is indeed absent in Sugden's approach. Second, we show that our concept is valid as an 'internal' critique of Sugden's, i.e. it is firmly embedded in the contractarian agenda. Finally, we briefly suggest some theoretical and policy implications.

We stipulate that most real-world agents have a desire to maintain their 'opportunity to learn' which we define as the set of potential

[18] There is a behavioural economics argument that may point in the opposite direction: Many real-world agents seem to underestimate the degree to which their preferences will change in future periods – this is referred to as 'projection bias' (e.g. Loewenstein and Angner 2003). Sugden does not adduce this piece of empirical evidence to support his case.

preferences that the agent has the capacity to learn, if he so chooses,[19] with 'learning' to be understood as any voluntary and cumulative acquisition of new preferences.[20] What motivates that desire? Let's assume an individual that wishes to optimize her opportunities over time, i.e. she acknowledges the fact that there may be a trade-off between short-term and long-term opportunity sets (see section 3, above). We stipulate that she also has an intuitive knowledge that unconditional endorsement of any preference that may come along in future periods may jeopardize that very goal: Maximizing opportunities, one may end up facing choice overload, which discourages choice, thus reducing the set of opportunities that are effectively available in future periods. In order to prevent self-defeating dynamics like this, the agent has to make sure to avoid ending up overwhelmed. Put differently, she has to make sure that her capacity to try out new experiences (such as goods or services) and, hence, her set of 'opportunities to learn' as defined above, is not reduced. Thus, the agent may wish to control her exposure to opportunities by engaging in some self-commitment, including self-constraint. Hence, while in Sugden's concept, the set of responses to anticipated preference change contains just one element (viz. unconditional endorsement of any new preference that may come along), our concept allows it to contain a variety of elements.

Given our definition of 'opportunity to learn', we postulate the following: *It is good that each person is free to maintain their set of opportunities to learn. In order to do so, they should be free to influence the way their own preferences develop over time (which includes being free to refrain from doing so) by making use of any kind of self-commitment devices they voluntarily choose, provided (i) the choice of devices is not irreversible, and (ii) others are not harmed in the process.*[21]

We suggest the criterion of 'opportunity to learn' as the proper yardstick to judge real-world institutional arrangements. When assessing them, we should ask: 'Do they allow individuals to try out and learn new preferences?' rather than 'Do they allow individuals to satisfy whatever preferences they happen to have?'

4.1 Yet another factor motivating self-constraint

Apart from avoiding choice overload and maintaining subjective coherence, individuals may wish to control their own exposure to

[19] Compare Sugden's definition of 'opportunity' as 'something [an individual] has the power to bring about, if he so chooses' (e.g. Sugden 2010: 49).

[20] 'Cumulative' here being meant to exclude phenomena such as preference reversal (Grether and Plott 1979) and mood-dependent switches in risk preferences (as in Sugden 2006b, 2007; see below).

[21] To illustrate, voluntarily selling oneself into slavery would be considered an irreversible act of self-commitment.

opportunities for a third reason that also presupposes some intuition (on their part) on the nature of preference dynamics. In other words, they are assumed to be aware of the fact that the space of possible preference paths is not perfectly open, but structured: Most preferences are path-dependent, some are habit-forming or even addictive. An intertemporal trade-off between opportunity sets not only appears when choice overload looms, but also when individuals learn preferences for products with certain (e.g. addictive) characteristics. While the former leads to choice abstention, the latter reduces future opportunity sets by causing 'blindness'.

In order to clarify this possibility, let us introduce the notion of 'perceived opportunity', meaning, in slight variation of Sugden's definition of the term 'opportunity' (see Sugden 2010: 49), something that the agent, from her subjective viewpoint, *perceives to have* – or: is aware of having – the power to bring about, if she so chooses. In most cases, real (or objective) and perceived opportunity will be roughly identical. Consider, though, being addicted to some kind of behaviour (such as drug consumption or gambling). Given the peculiar satiation characteristics of addictive consumption activities (see e.g. Witt 2001, 2010), agents may be 'trapped' in behavioural dynamics that systematically frustrate their attempts to satisfy their preferences in a more than transient way, thus incurring losses in terms of well-being. What is more important in our contractarian context, though, is the possibility that these agents 'turn blind' toward products and activities that are unrelated to their addiction, i.e. they suffer a loss in terms of (perceived) opportunities. While they may still be nominally free to pursue these alternative paths of preference acquisition, their perceived opportunity to do so can be impaired. Analogous reasoning may apply to the much-discussed issue of 'status races' (Frank 2008).[22] These are cases where a loss in perceived opportunities effectively reduces the agent's opportunities to learn. Anticipating this, agents may wish to limit their original opportunity sets by using tools of self-commitment, including self-constraint.

4.2 Isn't preference change already included in Sugden's account?

One might object at this point that Sugden's 'opportunity' concept does take sufficient account of preference change. Then, a dynamic variant such as the one we are proposing would be unnecessary. This, however, is not the case.

[22] Schubert and Cordes (2013) present a formal evolutionary model of these self-defeating preference dynamics.

At first sight, Sugden's concept seems well-tailored for a world in which preferences are not rock-solid and 'given', but unstable and evolving. For instance, it is supposed to allow for preferences to be 'unknown or undecided' (Sugden 2010: 50), and to be formed in the very process of interaction on markets for private goods (Sugden 2008a: 230). This fits nicely with the strategy to conceptualize preferences as pre-rational 'passions' rather than reason-based comparative evaluations. His 'opportunity' concept can be applied under these theoretically harsh conditions, because it addresses people's 'potential' rather than actual preferences (Sugden 2003: 791).

There is a flip side to this strategy, though. Conceptualizing preferences as 'passions' implies that theoretically, nothing further can be said about (i) their development and (ii) the way real-world individuals respond to this development. Passions escape theorizing, they are 'just there'. In fact, however, we can establish testable hypotheses about the driving forces of individual preferences and the cumulative and irreversible (i.e. genuine 'learning') nature of preference change (see e.g. Witt 2001, 2010). Tellingly, the only instant of something like systematic preference change in Sugden's work under discussion here is in Sugden (2006b, 2007), where – referring to Isen (1999) – he models an agent's (risk) preferences as being mood-dependent: they switch from one state to the other in a systematic, yet non-cumulative way, thus merely reflecting the instability of people's preferences.[23] We argue that this only captures a small part of preference change (see also Elster 1982).

As to people's diverse ways to respond to anticipated preference change, Loewenstein and Angner (2003: 376ff.), e.g., adduce evidence that agents' decisions about whether or not to indulge in future preferences (which translate into their judgements on opportunity) depend both on the source of preference change – such as 'visceral' or habit – and on the expected accuracy of predicting it. Relatedly, Senik (2008) shows that people are not indifferent between alternative temporal profiles of their future preference paths. Arguably, consideration of these issues marks a progress over simply claiming, ad hoc, that 'most' individuals will unconditionally wish to indulge in their future expected preferences.

Note that the possibility to theorize – at least within limits – about both preference change and agents' response to it does emphatically not imply that agents somehow end up being obliged to 'justify' their personal preferences to anyone. It seems that it is mainly worries about implications of that sort that drive Sugden's plea for conceptualizing preferences as pre-rational passions.

[23] Note also the emphasis Sugden (1998b: 42–43) puts on the *stochastic* nature of preference change.

4.3 Is this still a contractarian account?

One may object that the 'opportunity to learn' concept that we are proposing lacks sound contractarian credentials. This, however, is wrong for the following reason. In Sugden's view, a contractarian perspective implies, above all else, that the normative economist addresses citizens 'as they are', which presupposes (i) abstaining from the imposition of any specific theory of the good, and (ii) aiming at 'credible moral psychology' when modelling human behaviour: Real-world individuals, rather than some ethical observer, have to buy the standard of value that is being proposed.

As we already have discussed point (ii), let us focus on point (i): The normative concept that we propose does not aim at establishing some external standpoint from which the actual or potential preferences of individuals ought to be judged and possibly overridden. It does not rely on a 'reason to value' formula that is supposedly substantiated in some democratic process, reflecting a particular theory of well-being. The 'opportunity to learn' criterion simply points to the need to establish the institutional conditions that are required to maintain people's capacity to engage in the ongoing learning of new preferences. As we have seen, two of our three arguments – viz. the 'overwhelming choice' and the 'perceived opportunity' argument – do not depend on any *substantive* account of human well-being, i.e. they are valid even in a world where agents care about nothing but their opportunity sets. Thus, even for those agents a case can be made for the demand for self-commitment, including self-constraint.

4.4 Some implications

The shift in emphasis from 'opportunity' to 'opportunity to learn' opens up interesting new perspectives. Theoretically, the continuing person now appears as a 'locus of learning'. She would be concerned, at any given period of time, with having the opportunity to influence the way her future selves' preferences develop. Methodologically, this refocusing involves a shift toward a motivation-theoretic perspective that looks at *how sequences of preferences are interconnected through time* – rather than the decision-theoretic focus on isolated choices. We submit that this perspective is particularly adequate for a continuing person, understood as a composition of all her past, present and future 'time-slice selves', as Sugden puts it.

Hence, the appropriate question to ask is not whether individuals, when faced with the prospect of incoherent preferences, will *endorse* (Sugden's view) or *disavow* (the received view in normative behavioural economics) these preferences. Rather, the question is whether voluntarily chosen institutional arrangements are in place ensuring that the individuals' capacities to engage in the ongoing learning of new preferences are not jeopardized. Put differently, neither the present ('impulsive') nor any future ('planning') self ought to override the other self's substantive preferences; rather, our 'locus of learning' view of the responsible person implies that the continuing agent is interested, first and foremost, in making sure that all her selves be able to enjoy the opportunity to try out new preferences over time. Depending on the circumstances, either indulging in spontaneous preferences or listening to one's planning self may be better suited to achieve that goal.

Finally, two brief remarks on policy implications are in order. At the aggregate level, a social contract may include 'libertarian paternalist' devices that can support the agents' own attempts at self-commitment. Consider the introduction of mandatory 'cooling-off' periods for online purchases. In Sugden's view, this exemplifies an *apparent* restriction of individual freedom only (as opposed to a 'genuine' one), since it in fact enlarges the freedom of the individual, at least in her role as a consumer (see Sugden 2008b: 320). According to Sugden (2008b: 320), Mill's well-known prohibition of selling oneself into slavery is another case in point. In our view, though, the case for merely 'apparent' restrictions of this kind can only be consistently made on the grounds of our 'opportunity to learn' criterion: These restrictions can then be seen as institutional safeguards that help maintain an agent's capacity to engage in preference learning over time. In contrast, if one takes seriously Sugden's opportunity criterion, no 'responsible' agent (in the sense defined by him) could accept restrictions of this kind, since they curtail the opportunity set of his 'acting' self at the moment of choice. Weighing an increase in future opportunities against the immediate reduction of opportunities would imply according an active role to the planning self, something that Sugden's 'responsible person' is supposed to reject.

The general question one was to ask in thinking about policy implications would be: 'Would the individual be able, given her material circumstances, to acquire such-and-such preference?'[24] To illustrate, consider the way Sugden deals with the well-known 'adaptation problem' brought up by Sen (1988: 45–46). Sugden (2006a) focuses on the 'hopeless beggar's' nominal freedom to 'walk away', once he has 'changed his mind' (read: his preferences). This however hardly captures the normative intuition that made Sen worry about the case. We submit that it would be far more plausible to diagnose the beggar as having lost the *opportunity to learn*, where this presupposes the ability to *perceive* the opportunities available. Analogously, instead of asking 'What if the slave would cease to be contented ...' (Sugden 2006a: 46), we suggest to ask 'What if the

[24] Compare this to Sugden's guiding question: '*Were the individual to have* such-and-such a preference, would she be able to satisfy it?' (Sugden 2010: 50, italics in the original).

slave *cannot, on his own*, cease to be contented ...' Despite their (assumed) demonstrated happiness, the beggar and the slave have adapted to their situation by what would commonly be diagnosed as *resignation*.

5. CONCLUDING REMARKS

The present paper has argued that the 'opportunity criterion' proposed by Robert Sugden suffers from the fact that its microfoundations lack empirical plausibility. Specifically, the assumption that 'most' agents wish to maximize their opportunity sets is the Achilles heel of his account. Hence, Sugden's specification of contractarian interest is inadequate. We have suggested an alternative criterion that can be understood as a dynamic variant of Sugden's concept, in that it stresses people's ongoing opportunity to learn new preferences.

In light of our criterion, the market is particularly valuable from the perspective of individuals who do not yet have well-formed preferences (or possibly even no preferences at all) and *who still have to acquire them*. Agents then value markets to the extent that they provide the chance to do so and to try out new preferences over time. In order to achieve this, tools of self-commitment, including self-constraint, may be required that the market does not necessarily provide. Importantly, the mere fact that preferences are incoherent does not justify paternalistic intervention. We thus reject the claim made by some libertarian paternalists that the fact that social context shapes preferences 'undermines' the belief that agents should be free to choose and to make their own mistakes. Libertarian-paternalistic interventions are, however, legitimate in those (probably rare) cases where the acquisition of new preferences engenders learning dynamics that are likely to result in a loss of perceived opportunities, for that amounts to a loss of opportunities to learn over the long term; analogous reasoning applies to the introduction of mandatory 'cooling-off' periods.[25] Contrary to Sugden's view, real-world agents are unlikely to unconditionally endorse any preferences they may acquire in the future. What truly matters, in our view, is not the chance to satisfy whatever preference one may end up having in future periods, but, first and foremost, the chance to manage and develop one's preferences.

ACKNOWLEDGEMENTS

For helpful comments and discussions I would like to thank seminar participants at the universities of Freiburg, Utrecht and Witten-Herdecke, as well as Martin Binder, Malte Dold, Leonhard Lades, Maurizio Pugno,

[25] Neglecting this issue may even have significant distributive implications: As Beaulier and Caplan (2007) point out, the poor tend to be more susceptible to this kind of dysfunctional learning dynamics than the rich.

Johann Raphael, Robert Sugden, Martin van Hees, Ulrich Witt and an anonymous reviewer. All remaining errors and omissions are mine.

REFERENCES

Akerlof, G. A. and W. T. Dickens. 1982. The economic consequences of cognitive dissonance. *American Economic Review* 72: 307–319.

Arrow, K. 1995. A note on freedom and flexibility. In *Choice, Welfare and Development – A Festschrift in Honour of Amartya K. Sen*, ed. K. Basu, P. Pattanaik and K. Suzumura, 7–16. Oxford: Oxford University Press.

Beaulier, S. and B. Caplan. 2007. Behavioral economics and perverse effects of the welfare state. *Kyklos* 60: 485–507.

Berg, N. and G. Gigerenzer. 2007. Psychology implies paternalism? Bounded rationality may reduce the rationale to regulate risk-taking. *Social Choice and Welfare* 28: 337–359.

Bernheim, D. B. and A. Rangel. 2009. Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics* 124: 51–104.

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: 1211–1254.

Elster, J. 1982. Sour grapes – utilitarianism and the genesis of wants. In *Utilitarianism and Beyond*, ed. A. K. Sen and B. A. Williams, 219–238. Cambridge: Cambridge University Press.

Frank, R. H. 2008. Should public policy respond to positional externalities? *Journal of Public Economics* 92: 1777–1786.

Frankfurt, H. G. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy* 68: 5–20.

George, D. 2004. *Preference Pollution*. Ann Arbor, MI: University of Michigan Press.

George, D. 2006. Moral implications of preference change. *Society* 43: 33–38.

Grether, D. M. and C. R. Plott. 1979. Economic theory of choice and the preference reversal phenomenon. *American Economic Review* 69: 623–638.

Hayek, F. A. 1989. Order – with or without design? London: Centre for Research into Communist Economies.

Hoeffler, S. and D. Ariely. 1999. Constructing stable preferences: a look into dimensions of experience and their impact on preference stability. *Journal of Consumer Psychology* 8: 113–139.

Isen, A. M. 1999. Positive affect. In *Handbook of Cognition and Emotion*, ed. T. Dalgleish and M. J. Power, 521–539. Chichester: Wiley.

Iyengar, S. and M. Lepper. 2000. When choice is demotivating: can one desire too much of a good thing? *Journal of Personality and Social Psychology* 79: 995–1006.

Loewenstein, G. 1999. Willpower: a decision-theorist's perspective. *Law and Philosophy* 19: 51–76.

Loewenstein, G. and E. Angner. 2003. Predicting and indulging changing preferences. In *Time and Decision*, ed. G. Loewenstein, D. Read and R. Baumeister. New York, NY: Russell Sage, 351–391.

McQuillin, B. and R. Sugden. 2012. How the market responds to dynamically inconsistent preferences. *Social Choice and Welfare* 38: 617–634.

O'Donoghue, T. and M. Rabin. 1999. Doing it now or later. *American Economic Review* 89: 103–124.

Read, D. 2006. Which side are you on? The ethics of self-command. *Journal of Economic Psychology* 27: 681–693.

Roemer, J. E. 1998. *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Rothenberg, J. 1962. Consumers' sovereignty revisited and the hospitability of freedom of choice. *American Economic Review – Papers and Proceedings* 52: 269–283.

Sally, D. 2000. I, too, sail past: Odysseus and the logic of self-control. *Kyklos* 53: 173–200.

Schelling, T. C. 1978. Egonomics or the art of self-management. *American Economic Review* 78: 290–294.

Schelling, T. 1984. Self-command in practice, in policy and in a theory of rational choice. *American Economic Review* 74: 1–11.

Schubert, C. 2012. Pursuing happiness. *Kyklos* 65: 245–261.

Schubert, C. and C. Cordes. 2013. Role models that make you unhappy: light paternalism, social learning and welfare. *Journal of Institutional Economics* 9: 131–159.

Schwartz, B. 2004. *The Paradox of Choice*. New York, NY: HarperCollins.

Sen, A. K. 1988. *On Ethics and Economics*. Oxford: Basil Blackwell.

Sen, A. K. 1992. *Inequality Reexamined*. Cambridge, MA: Harvard University Press.

Senik, C. 2008. Is man doomed to progress? *Journal of Economic Behavior and Organization* 68: 140–152.

Shin, J. and D. Ariely. 2004. Keeping doors open: the effect of unavailability on incentives to keep options viable. *Management Science* 50: 575–586.

Sugden, R. 1989. Maximizing social welfare: is it the government's business? In *The good Polity*, ed. A. Hamlin and P. Pettit. Oxford: Basil Blackwell, 69–86.

Sugden, R. 1998a. The metric of opportunity. *Economics and Philosophy* 14: 307–337.

Sugden, R. 1998b. Measuring opportunity: toward a contractarian measure of individual interest. *Social Philosophy and Policy* 15: 34–60.

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

Sugden, R. 2004a. The opportunity criterion: consumer sovereignty without the assumption of coherent preferences. *American Economic Review* 94: 1014–1033.

Sugden, R. 2004b. Living with unfairness: The limits of equal opportunity in a market economy, *Social Choice and Welfare* 22: 211–326.

Sugden, R. 2006a. What we desire, what we have reason to desire, whatever we might desire: Mill and Sen on the value of opportunity. *Utilitas* 18: 33–51.

Sugden, R. 2006b. Taking unconsidered preferences seriously. In *Preferences and Well-Being*, ed. S. Olsaretti. Cambridge: Cambridge University Press, 209–232.

Sugden, R. 2007. The value of opportunities over time when preferences are unstable. *Social Choice and Welfare* 29: 665–682.

Sugden, R. 2008a. Why incoherent preferences do not justify paternalism. *Constitutional Political Economy* 19: 226–248.

Sugden, R. 2008b. Capability, happiness and opportunity. In *Capabilities and Happiness*, ed. L. Bruni, F. Comim and M. Pugno, 299–322. New York, NY: Oxford University Press.

Sugden, R. 2009. Market simulation and the provision of public goods: a non-paternalistic response to anomalies in environmental valuation. *Journal of Environmental Economics and Management* 57: 87–103.

Sugden, R. 2010. Opportunity as mutual advantage. *Economics and Philosophy* 26: 47–68.

Sugden, R. 2013. The behavioural economist and the social planner: to whom should behavioural welfare economics be addressed? *Inquiry* 56: 519–538.

Sugden, R. 2014. Characterising competitive equilibrium in terms of opportunity. Working Paper, University of East Anglia.

Sunstein, C. R. and R. H. Thaler. 2003. Libertarian paternalism is not an oxymoron. *University of Chicago Law Review* 70: 1159–1202.

Thaler, R. H. and C.R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Witt, U. 2001. Learning to consume – A Theory of Wants and the Growth of Demand. *Journal of Evolutionary Economics* 11: 23–36.

Witt, U. 2010. Economic behavior – evolutionary vs behavioral perspectives. Papers on Economics and Evolution # 1017, Max Planck Institute of Economics, Jena, Germany.

Yaari, M. E. 1977. Endogenous changes in tastes: a philosophical discussion. *Erkenntnis* 11: 157–196.

BIOGRAPHICAL INFORMATION

Christian Schubert is Visiting Professor of Economic Ethics at the University of Kassel, Germany. His current scholarship focuses on the implications of behavioural economics for welfare economics and economic ethics and the ethical assessment of concepts such as 'Libertarian Paternalism'.