International Review of Economics (2021) 68:67–88
https://doi.org/10.1007/s12232-020-00361-x

RESEARCH ARTICLE

Informed preference consequentialism, contractarianism and libertarian paternalism: on Harsanyi, Rawls and Robert Sugden's The Community of Advantage

Mozaffar Qizilbash[1]

Received: 8 June 2020 / Accepted: 24 September 2020 / Published online: 19 October 2020

Abstract
Robert Sugden abandons certain central tenets of traditional welfare economics and recommends a contractarian alternative. He rejects 'Libertarian Paternalism' (LP) and related 'paternalistic' proposals. The seeds of 'paternalism' inspired by the findings of behavioural economics can be found in informed preference views associated with J.S. Mill and John Harsanyi. Nonetheless, those who endorse a combination of the informed preference view of welfare, consequentialism and welfarism—'informed preference consequentialists'—have good reasons to resist the agenda of LP. John Rawls adopts a variation of the informed preference view. Contracting parties in his theory accept 'paternalistic principles'. Sugden's claim that contractarians cannot be 'paternalists' does not generalise to all contractarian theories. Sugden's and Rawls' contractarian positions are in important respects different.

Keywords Consequentialism · Contractarianism · Welfare · Paternalism · Behavioural economics

JEL Classification A12 · D60 · D61 · D63 · D90

1 Introduction

In The Community of Advantage and related works, Sugden (1989, 2018a, p. 17) rejects some of the central tenets of the utilitarian heritage of welfare economics. These include two views which utilitarians endorse: (1) *consequentialism*—the view that the right action, rule or motive is one which leads to the outcome or state of affairs which is best (or no worse than any other) and; (2) *welfarism*—the view that the relative goodness of different outcomes or states of affairs depends

✉ Mozaffar Qizilbash
Mozaffar.qizilbash@york.ac.uk

[1] Department of Economics and Related Studies, University of York, York YO10 5DD, England

only on the levels of individual welfare in those outcomes or states (see Sen 1979, pp. 463–468). He also abandons the view that a person's advantage should be evaluated in terms of the satisfaction of her preferences. Sugden (2018a, p. 17) hopes to 'encourage the reader to wonder whether there is merit in alternative approaches'. Sugden (2018a, p. 19) suggests that traditional welfare economics is wedded to a peculiar 'view from nowhere'—a term he borrows from Nagel (1986)—which involves taking the imaginary perspective of an 'impartially benevolent spectator', and which 'attempts to filter out one's private interests and biases'. Consequentialism is sometimes defined so that it takes precisely this sort of view. For example, Samuel Sheffler writes that: '[c]onsequentialism in its purest and simplest form is a moral doctrine which says that the right act in any given situation is the one that will produce the best overall outcome, as judged from an impersonal standpoint which gives equal weight to the interests of everyone' (Sheffler 1988, p. 1). Sugden (2018a, p. 20) also argues that a 'view from nowhere' has implicitly been adopted by those, like Cass Sunstein and Richard Thaler, who advance various forms of 'paternalism' in extending normative economics in the light of the findings of Behavioural Economics (BE). Sugden rejects this approach and advances a contractarian alternative which grounds morals on agreement and adopts an opportunity-based view of advantage. He argues that contractarians cannot be 'paternalists' (Sugden 2018a, pp. 42–45).

The Community of Advantage defends the liberal tradition in normative economics and recommends contractarianism to liberals who are troubled by the rise in 'paternalistic' proposals in normative economics and public policy. In this paper, I focus on only one of these proposals: Sunstein and Thaler's libertarian paternalism (LP, for short). In its attempt to 'extend' the logic of traditional welfare economics, LP implicitly endorses a version of the informed or rational preference view according to which welfare is constituted by the satisfaction of informed or rational preferences. J.S. Mill arguably articulated a version of this view, and variants of it were also advanced by John Harsanyi. Since Mill and Harsanyi were both utilitarians, they accepted both consequentialism and welfarism. One question which might be posed by readers of Sugden's book is whether a combination of welfarism, consequentialism and the informed preference view—which, when combined, I term 'informed preference consequentialism'—necessarily encourages 'paternalism' based on the findings of BE. In this paper, I give a negative answer to this question. I argue, nonetheless, that there are aspects of Mill's and Harsanyi's views which contain the seeds of the relevant 'paternalistic' proposals. These aspects of Mill's and Harsanyi's views are arguably also present in John Rawls' original statement of his contractarian theory of justice, which endorses a version of the informed preference view. This suggests that Rawls' contractarian account may be compatible with LP. I examine Sugden's claim that a 'contractarian cannot be a paternalist' and the relationship between Rawls' and Sugden's views in this context.

The paper is structured as follows: Sect. 2 introduces the informed preference view in the works of Mill and Harsanyi and the forms of 'paternalism' they may encourage or resist; in Sect. 3, I argue that there are good arguments which suggest that informed preference consequentialists should reject the agenda of LP; I turn to

John Rawls' contractarian theory and Sugden's claim that contractarians cannot be 'paternalists' in Sect. 4 and Sect. 5 concludes.

2 Informed preference views and 'paternalism'

The informed preference view of welfare can be traced to J.S. Mill's writings. Following his statement of utilitarianism, Mill introduces the distinction between qualities of pleasures:

> If I am asked, what I mean by difference in quality of pleasures, or what makes one pleasure more valuable than another, merely as a pleasure, except its being greater in amount, there is but one answer. Of two pleasures, if there is but one to which all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure. (Mill 1962, p. 257).

Those who have 'experience of both' are 'competent judges'. Crisp (1997, p. 29) notes that the test of the judgement of the 'competent judges' is a version of the 'informed preference test' which was more explicitly and fully articulated in utilitarian thought by Henry Sidgwick (1881) amongst others (see Griffin 1986; Harsanyi 1981, 1995, 1997; see also Qizilbash 1998, 2006 inter alia).

John Harsanyi's statements of utilitarianism articulate different versions of the requirement for a preference to count as 'rational' or 'informed'. In one, 'true preferences' are preferences someone 'would have if he had all the relevant factual information, always reasoned with the greatest possible care, and were in a state of mind most conducive to rational choice' (Harsanyi 1981, p. 55). These are contrasted with his actual preferences which are manifested in behaviour and which may fall short of 'true preferences' because of 'erroneous factual beliefs, careless logical analysis and strong emotions which hinder rational decision making'. In a later articulation of his view, 'informed preferences' are 'the hypothetical preferences he would have if he had all the relevant information and had made full use of the information' (Harsanyi 1997, p. 133). In advancing his moral theory, Harsanyi (1981, pp. 44–48) takes a 'view from nowhere' in as much as he models the impartial standpoint in terms of an imagined situation where people do not know what position they will take in society and have an equal probability of being in any social position. Harsanyi argues that from this standpoint, rational agents make a choice which is 'independent of morally irrelevant selfish considerations' (Harsanyi 1981, p. 45). He suggests that agents who maximise expected utility will choose the society with the highest average welfare. And he argues in favour of a form of utilitarianism which recommends the rules which maximise average utility (Harsanyi 1981, pp. 56–60).

Since, on Harsanyi's view, the satisfaction of actual preferences may not constitute welfare, it may be problematic for him to make the standard case for the market. That case is usually encapsulated in the first theorem of welfare economics which states that, given standard assumptions, the competitive market

is 'Pareto efficient'—in the sense that no reallocation of commodities can make one consumer better off without rendering another worse off. Because it is a welfare theorem, the terms 'better off' and 'worse off' here are usually understood in terms of welfare. This theorem is central to the defence of the market in economics, and is regularly interpreted in terms of Adam Smith's notion of the 'invisible hand' (e.g. Mas-Colell et al. 1995, p. 327) because the maximising choices of individuals have as an unintended consequence an outcome which is 'good' at the social level, at least in as much as it is efficient. In its standard form, this theorem supports a central pillar of traditional welfare economics, which I refer to as 'Qualified Market Efficiency' (or QME, for short) and which claims that: under certain conditions, the (unintended) consequence of the maximising choices of consumers and producers, given their constraints, is a (competitive or) market equilibrium which is Pareto efficient in terms of welfare. If the relevant conditions are not met, there is 'market failure' and a prima facie case for government interference.

QME retains the focus on consequences and welfare in traditional welfare economics. It understands welfare in terms of the satisfaction of actual preferences, rather than true or informed preferences of the sort that Harsanyi had in mind.[1] To the degree that actual and informed or true preferences diverge, on the informed preference view, the first welfare theorem only supports a case for Pareto efficiency of the market in terms of actual preference satisfaction (see Sen 1993; Qizilbash 2018). Indeed, QME would (on this view), at least in principle, justify government intervention on the grounds of 'market failure' when people's behaviour does not track their true or informed preferences. As we shall see, Harsanyi's position here contains the seeds of the case for 'paternalistic' interventions to correct 'behavioural market failures' which has been advanced in the recent literature on BE.

Does Harsanyi's account of welfare necessarily justify interventions in people's lives of the sort that 'anti-paternalists' find objectionable? To make a start on answering this question, we need a definition of 'paternalism'. The most widely used definition of it is Gerald Dworkin's. On this definition, 'paternalism' is 'interference with a person's liberty of action justified by reasons referring exclusively to the welfare, good, happiness, needs, interests or values of the person being coerced' (Dworkin 1971, p. 108). Because on this definition, 'paternalistic' interference must involve coercion, I refer to it as 'hard paternalism'. And Mill clearly objected to this when he advanced his 'harm principle'. He wrote that:

> [T]he only purpose for which power can rightfully be exerted over any member of a civilised community, against his will, is to prevent harm to others. His

[1] It can be argued that neo-classical economists do sometimes assume that the producers and consumers whose choices are relevant to QME have better information than anyone else about specific domains. In this distinct sense, it is assumed that, in the case for the market, consumers have informed preferences. See, for example, Arrow (1983, pp. 200–201). Indeed, it may be for this reason that it is sometimes supposed that economists more generally hold the informed preference view. On this, see also Hausman (2012, pp. 83–87). Nonetheless, the requirement for a preference to be informed is here quite distinct from the sort Harsanyi had in mind.

> own good, either physical or moral, is not sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others it would be wise, or even right (Mill 1962, p. 135).

While Mill rejects 'hard paternalism' in general, he allows for certain exceptional cases where coercion might be justified, including cases where someone does not have relevant information. One well-known exception involves someone approaching an unsafe bridge who may be unaware that it is not safe. Mill suggests that anyone who sees someone doing this might 'seize him and turn him back', since while 'liberty consists in doing what one desires … he does not desire to fall into the river'. Mill qualifies his remarks because there may be some uncertainty about the situation (in this example). In general, he suggests that people should 'only be warned of the danger; not forcibly prevented from exposing' themselves to it. But he still admits exceptions where coercion might be appropriate: the cases of children, and those who are 'delirious, or in some state of excitement or absorption incompatible with the full use of the reflecting faculty' (Mill 1962, pp. 228–229). Mill here opens the door to coercive intervention in exceptional cases if desires are irrational.

Does Harsanyi permit coercive interference in people's choices when they do not act on their informed or true preferences? Harsanyi (1997, p. 134) asks himself whether 'paternalism' is justified when people act on 'mistaken preferences', preferences which are not consistent with their informed preferences. In particular, he asks himself '[t]o what extent should our society follow a liberal policy, permitting people to "make their own mistakes", and to what extent should it follow a paternalistic policy, trying to prevent people from self-damaging behavior?' His answer to this question is that: 'in a democratic society, positive paternalism, which would try coercively to prevent self-damaging behaviour, can be justified only in cases where such behaviour would inflict utterly intolerable damage on the agent…' (Harsanyi 1997, p. 139). Harsanyi thus follows Mill in opposing hard paternalism in all but exceptional cases.

Since Harsanyi does not expand much on his views about 'paternalism', it is worth considering one argument which suggests that the informed preference view would not necessarily justify coercive interference when people act on 'mistaken preferences'. To develop this argument, consider James Griffin's informed desire view which influenced Harsanyi's views (see, in particular, Harsanyi 1995, 1997). On Griffin's account, informed desires are desires 'formed by an appreciation of the nature of the object' (Griffin 1986, p. 14). The objects of informed desire are the things that make a characteristically human life go better, or 'prudential values'. Griffin advances a list of these. One value on the list is 'autonomy'. He writes:

> Choosing one's own way through life, making something out of it according to one's own lights, is at the heart of what it is to lead a human existence. And we value what makes life human, over and above what makes it happy. What makes life 'human' in the distinctly normative sense it has here, is not a simple thing. The systematic way to understand its complexities is to understand the complexities of 'agency'. One component of agency is deciding for oneself. Even if I constantly made a mess of my life, even if you could do better if you

took charge, I would not let you do it. Autonomy has a value of its own. (Griffin 1986, p. 67).

On Griffin's account allowing people to make their own mistakes (even when others could do better) is part of what upholding the value of autonomy requires, since deciding for oneself is a component of that value (see also Griffin 2008, pp. 150–157). As a consequence, interfering with people's lives to avoid such errors may be objectionable. This is the first argument from autonomy. If Harsanyi accepted this argument, it would help to explain why he believed that people should be permitted, except in rare cases, to act on 'mistaken preferences'.[2]

3 Libertarian paternalism and informed preference consequentialism

The findings of BE and psychology have inspired various forms of interventionist public policy. Some of the proposals which have emerged suggest that coercive measures are justified to further people's welfare (see Camerer et al. 2003; Conly 2013). By contrast, Cass Sunstein and Richard Thaler class interference in people's choices as 'paternalistic' as long as it promotes their welfare; but the relevant interventions are not coercive to the degree that they do not block choice. Such interference is a form of 'soft paternalism'. While Sunstein and Thaler (2003, p. 1163) avoid taking any contentious view of the components of welfare, they implicitly adopt a variation of the informed or rational preference view of welfare (see Sugden 2008a, p. 232; Qizilbash 2012). They tell us that 'in some cases people make inferior decisions in terms of their own welfare—decisions which they would change if they had complete information, unlimited cognitive abilities, and no lack of self-control' (Sunstein and Thaler 2005, p. 176). They argue that in some such cases—involving status quo bias, framing, self-control, myopia and so on—BE and psychology teach us that people systematically make choices which are either not in their best interests or irrational. To take a well-known example, in a cafeteria, people are more attracted to food which is presented earlier in the queue and at eye level. If unhealthy food is presented early in the queue and at eye level, then people will make less healthy choices than they would if it was not so prominently displayed. Of course, if a 'paternalist' planner or 'choice architect'—were to reframe the choice to make healthier options more attractive, then she has done nothing coercive, and the intervention does not block choice. It is only, on Sunstein and Thaler's view, a form of 'paternalism' to the degree that it involves interference with a view to improving welfare. Furthermore, the claim is that such interference improves people's decision-making and welfare as judged by themselves. Nonetheless, some argue that if the intervention

[2] There is reason to think that Harsanyi would have endorsed this position. In his later works, he followed Griffin in advancing a list of prudential values, one of which was 'freedom to control our own lives' (Harsanyi 1995, p. 323).

shapes (or manipulates) choice—by pushing agents in one direction rather than another, it undermines autonomy, where autonomy is understood in terms of a person's control over her or his environment. When it does so, such interference arguably involves an objectionable form of 'paternalism' (see Hausman and Welch 2010, p. 128; Bovens 2008). It is (arguably) objectionable when it violates freedom in this way even if there is no coercion.[3] This is the second argument from autonomy.[4]

Sunstein and Thaler recommend several interventions which are justified by the failure of people to act according to informed or rational preferences. They implicitly take the standard of rationality to be that set by expected utility, or rational choice, theory. The behaviour of the rational agents of (neo-classical) economic theory is that of 'Econs', and those of us who fall short of that standard in predictable ways are 'Humans' (Thaler and Sunstein 2008, pp. 7–9). Interventions to improve the welfare of Humans—which are examples of what they term 'nudges'—are justified by predictable differences between Humans and Econs. In treating the standard for an 'informed' or 'rational' preference to be that set by rational choice theory (RCT) or economics, Thaler and Sunstein adopt a strong standard for 'informed preference' (see Qizilbash 2012). This point also comes out in an argument which Thaler and Sunstein make against 'anti-paternalists'. The argue (contentiously) that 'anti-paternalists' assume that 'almost all people, almost all of the time, make choices that are in their best interest or at least are better than the choices that would be made by someone else' (Thaler and Sunstein 2008, p. 10). They contest this assumption using an example involving a novice playing an experienced chess player, and they conclude that: 'so long as people are not choosing perfectly, some changes in the choice architecture could make their lives go better (as judged by their preferences, not those of some bureaucrat)' (Thaler and Sunstein 2008, p. 10). 'Choosing perfectly' sets a very high standard for rationality. This standard, in turn, can justify a very wide range of interventions. By contrast, Mill's implicit version of the informed preference test in Utilitarianism required only that the chooser is 'competent' in the sense that she has experience of the relevant available options. An experienced player may, by and large, make better choices than a novice, but there is no implication that she would 'choose perfectly'. Indeed, Mill's reference to the 'full use of the reflecting faculty' in On Liberty predates the emergence of modern RCT, and would not have required human beings to choose in a way that is consistent with its axioms.[5]

Next consider market failure. Sunstein (2014, p. 16) argues—like others (e.g. Camerer et al. 2003, p. 1218) in this literature—that '[t]he various empirical findings

[3] For this reason, Hausman (2018, p. 55) notes that the definition of 'paternalism' can be expanded to include 'interference with liberty *or autonomy* of the person whom the action aims to benefit'. See also Gerald Dworkin's revised definition of 'paternalism' in Dworkin (2020).

[4] While I have listed 'anti-paternalist' arguments which focus on the value of autonomy, there are also autonomy-based arguments for 'paternalism'. For an example, see Sunstein (1991).

[5] Advocates of LP might, nonetheless, argue that Mill would not have regarded libertarian paternalistic interventions as problematic, since, aside from rare exceptions, he objected to 'hard' rather than 'soft paternalism'.

allow us to identify a set of behavioural market failures, understood as market failures that complement the standard economic account and that stem from the human propensity to err'. This extension of the concept of market failure to cases where people's preferences diverge from their 'true' or 'informed' preferences follows the logic of Harsanyi's utilitarianism. Furthermore, while Harsanyi objects to 'positive paternalism' which would involve coercion except in exceptional circumstances, he does not rule out soft paternalism.

Should informed preference consequentialists endorse LP? There are reasons for them to hesitate before doing so. One is that the results of BE and psychology are as relevant to the choices of the Humans who are the targets of libertarian paternalistic interventions as to those of 'planners' who seek to construct the relevant 'choice architecture'. If so, it is implausible that the results of BE and psychology support the crucial claim that fallible third parties can choose or do any better than the Humans whose behaviour apparently falls short of that of Econs (see, in particular, Glaeser 2006). As a result, a world with interventions inspired by LP may be worse—or no better—than one without these. The case for these interventions is weaker still if there is scepticism about the reliability or robustness of the behavioural findings on the basis of which 'paternalism' might be recommended. The relevant interventions may, thus, not be accepted on a consequentialist calculus.

A further reason to hesitate has to do with the assumed benevolence of those designing the relevant interventions. 'Choice architects' may be not merely fallible, but also self-interested or malign. If, for example, the relevant policy makers wish to manipulate behaviour so as to pursue their own interests (e.g. to maximise the budgets of their departments), their behaviour would not even count as 'paternalistic' (on any plausible definition), because it would not be motivated with a view to promoting (other) people's welfare or interests. Interventions of this sort by 'planners' would thus undermine the claim that libertarian paternalists sometimes make that 'paternalism is inevitable' (e.g. Sunstein and Thaler 2005, p. 178; Sunstein 2014, p. 121).[6] There are thus good reasons for informed preference consequentialists to hesitate before endorsing the agenda of LP.

There are at least three 'anti-paternalist' arguments which emerge from this discussion and which derive from informed preference views:

(I) To the degree that the ability to make one's own 'mistakes' is part of what the value of autonomy requires, people's 'mistakes' do not, in general, justify interference in their choices;

(II) To the degree that interference based on the findings of BE shapes (or manipulates) people's choices, it may violate their autonomy, and may for this reason be objectionable; and.

(III) The standard that LP sets for preferences to be 'informed' or 'rational' is very demanding, and recommends interference that is not justified by a divergence

[6] Sometimes the claim that 'paternalism is inevitable' appears to be mistakenly conflated with or run together with the claim that 'choice architecture is inevitable'. See, for example, Sunstein (2014, p. 118 and 121). Yet one might well reject the first of these claims while accepting the latter.

of informed (or rational) and actual preferences on less demanding versions of the informed preference view.

In (I), there are quotation marks around 'mistakes' since—given (III)—failing to behave according to the axioms of RCT would not even count as a 'mistake' on some informed or rational preference views.[7] (I) and (II) are versions of the first and second arguments from autonomy. While I introduced the first of these in the context of Griffin's view and Harsanyi's rejection of 'positive paternalism', its logic extends to the interventions proposed by LP, since LP justifies interference on the basis of 'mistakes'.

There is another reason why Harsanyi might object to the full range of interventions favoured by LP. This relates to what Sunstein (2014, pp. 116–118) thinks of as a 'rule-consequentialist anti-paternalist' argument against LP. By way of clarification, in this context, 'rule-consequentialism' is any version of consequentialism which focuses on rules (rather than actions or motives).[8] This argument can be formulated as follows:

(IV) Given the wide range of areas where LP recommends interference to improve people's welfare and decision-making, the consequences of adopting any set of rules (social norms or moral code) which recommends such extensive interference would be worse (or no better) than the status quo.

There is more than one reason to accept (IV). One has already been discussed: 'choice architects' are as fallible as any other Human, and a world with their wide-ranging interference may well be worse (or no better) than one without it. But even if we set aside their fallibility, a world with so much interference in people's lives would arguably be worse than one without it. It appears to be the first of these reasons—which he associates with Glaeser (2006)[9] and—which Sunstein believes motivates the 'rule-consequentialist anti-paternalist' argument. Nonetheless, the wide range of interventions justified by a failure of people to 'choose perfectly' and to align 'perfect' and actual choice also supports (IV). Accepting any of (I)–(IV) may lead an informed preference consequentialist to reject what Sunstein (2014, p. 17) calls the 'First (and) only Law of Behaviorally Informed Regulation: *In the face of behavioural market failures, nudges are usually the best response, at least when there is no harm to others'*. However, for those who endorse all of (I)–(IV), the case for rejection is overwhelming.

Sunstein (2014, pp. 116–122) considers some of these 'anti-paternalistic' arguments and argues that, looking at specific cases, any strong presumption against interference must fail. In the case of objections involving the value of autonomy, he

[7] An example is Griffin's view, which allows for informed preferences which are non-transitive. See Griffin (1986, pp. 96–97).

[8] This sense of 'rule-consequentialist' identifies a class of moral theory, rather than a specific moral theory such as Brad Hooker's. See Hooker (2000).

[9] See Sunstein (2014, p. 180, notes 10 and 11).

argues that there is a risk of exaggerating their force, since consequentialists must sometimes trade-off autonomy against other values. I will not evaluate Sunstein's responses here. Rather I suggest that because Harsanyi was a rule-utilitarian—and so a rule-consequentialist—he may have accepted the 'rule-consequentialist anti-paternalist' argument which Sunstein contests.[10] Indeed, Harsanyi advanced a similar argument in defence of the value of personal choice. He argued that: 'it is better to live in a society which allows people a good deal of free choice in their personal lives and does not impose unacceptably burdensome restrictions on people's personal behavior' (Harsanyi 1995, p. 330). My guess is that Harsanyi would equally have resisted widespread interference on the basis of the results of BE, especially when this might undermine autonomy. So while the idea of a 'behavioural market failure' can be traced to the logic of Harsanyi's utilitarianism, it, by no means, follows that Harsanyi would have endorsed the agenda of LP. Indeed, the same can be said of informed preference consequentialists more generally. There are good reasons why informed preference consequentialists should reject the 'First Law of Behaviorally Informed Regulation' (henceforth, FLoBIR). Indeed, 'anti-paternalists' need not reject informed preference consequentialism on the grounds that it would commit them to the agenda of LP.[11]

4 Sugden's Contractarian 'Anti-Paternalism' and Rawls' Theory of Justice

Robert Sugden's argument against various forms of 'paternalism' which are inspired by BE is quite distinct from consequentialist 'anti-paternalist' arguments. He has argued that morality is not about maximising social welfare (Sugden 1989). Sugden holds a contractarian view, which sees morality in terms of an agreement between individuals. In The Community of Advantage Sugden follows James Buchanan's account of contractarianism. He cites a passage from Buchanan which runs:

> If politics is to be interpreted in any justificatory or legitimising sense without the introduction of supra-individual value norms, it must be modelled as a process within which individuals, with separate and potentially different inter-

[10] While this point holds for Harsanyi, it may not hold for Mill, since Mill's definition of utilitarianism focusses on actions: 'actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness' (Mill 1962, p. 257). On the question of whether Mill was an act- or a rule-utilitarian, see Crisp (1997, pp. 102–105 and 116–117).

[11] Another alternative might start from a reading of Amartya Sen's work—due to Siddiq Osmani (2018)—which suggests that various forms of supposed 'irrationality' discussed in BE would not count as forms of 'irrationality' on his account. This would imply that—on Sen's account—'paternalistic' interference on the basis of the findings of BE is not justified. Sen (2000; 2002) also endorses a form of consequentialism and an undemanding view of rationality. An 'anti-paternalist' consequentialist alternative to LP might emerge from this reading of Sen. Nonetheless, Sen would not qualify as an informed preference consequentialist, since he rejects 'welfarism' (see Sen 1979). Sugden (1981, 1985, 2006b, 2008b, 2010a, 2018a, pp. 24–28; 2020; see also Sugden 1978) has forcefully criticised aspects of Sen's position. See also Sen (2006) and Qizilbash (2011a, 2011b) *inter alia*.

> ests and values, interact for the purpose of securing individually valued benefits of co-operative effort. If this presupposition about the nature of politics is accepted, the ultimate model of politics is *contractarian*' (Buchanan 1986, p. 240; Sugden 2018a, p. 29).

As a result, Sugden's contractarian proposals are not addressed to planners—in the way that LP is—but to 'individuals together' (Sugden 2018a, p. 46). As Sugden (2018a, p. 48) puts it, if one were to ask why, on his contractarian view, 'paternalism is out of bounds? The answer is not that, all things considered, paternalism has undesirable consequences.' Sugden argues rather that a contractarian cannot be a 'paternalist' (Sugden 2018a, pp. 42–50), where 'paternalism' is defined—following Le Grand and New (2015, p. 23)—in terms of 'government intervention' which 'is intended to address a failure of judgement by the individual' and 'to further that individual's good' (Sugden 2018a, p. 46). Sugden's argument has force within his own contractarian account. Any proposal he advances is addressed to 'individuals as directors of their lives' (Sugden 2018a, p. 49). It is their judgements that prevail rather than those of policy makers or others who might decide for them. To this degree, a contractarian cannot be a 'paternalist'. There remains the question of whether Sugden's claim that a contractarian cannot be a 'paternalist' also holds in, and can be generalised to, other contractarian accounts.

To explore this question, I focus exclusively on one influential contractarian theory—the theory of 'justice as fairness' advanced in its original form by John Rawls in A Theory of Justice. As he sets out his theory, Rawls (1972, p. 11) notes that its 'guiding idea is that the principles of justice for the basic structure of society are the object of the original agreement'. By the 'basic structure' he has in mind, 'its major institutions' which include 'the political constitution and the principal economic and social arrangements' (Rawls 1972, p. 7). The theory involves the choice of principles of justice behind a 'veil of ignorance' in an 'original position' where 'no one knows his place in society, in the distribution of natural abilities, his intelligence, strength and the like'. Nor do individuals know their 'conception of the good or their special psychological propensities' (Rawls 1972, p. 12). In Rawls' theory, 'the original position of equality corresponds to the state of nature in the traditional theory of the social contract' (Rawls 1972, p. 12). It is in this position that principles of justice are chosen. Because the parties in the original position 'are similarly situated and no one is able to design principles to favour his particular conditions, the principles of justice are the result of a fair agreement or bargain' (Rawls 1972, p. 12). These principles also presuppose a conception of advantage and Rawls (1972, chapter VII) adopts a rational desire view. A person's advantage is understood in terms of those 'things every rational man is presumed to want' and which 'have a use whatever a person's rational plan of life' (Rawls 1972, p. 62). These 'things' are 'primary goods'. And Rawls is concerned with those primary goods which are at the 'disposition of society': the 'social primary goods'.

Rawls' theory is primarily concerned with principles of justice rather than welfare economics. In making some remarks about economic systems, he notes that his subject is 'the theory of justice and not economics, however elementary' (Rawls 1972, p. 265). However, he does engage with the question of 'how the

two principles work out as a conception of political economy, that is as standards to assess economic arrangements and policies, and their background institutions'. He then adds in parentheses that '[w]elfare economics is often defined in the same way. I do not use the name because the term "welfare" suggests that the implicit moral conception is utilitarian …' (Rawls 1972, pp. 258–259).

Rawls briefly discusses the market system, one of the chief merits of which, he believes, is efficiency. He describes the first welfare theorem as follows:

> Under certain conditions, competitive prices select the goods to be produced and allocate resources to their production in such a manner that there is no way to improve on either the choice of productive methods by firms, or the distribution of goods that arises from the purchases of households. There exists no rearrangement of the resulting economic configuration which makes one household better off (in view of preferences) without making another worse off. No mutually advantageous trades are possible … (Rawls 1972, pp. 271–272).

Rawls here follows orthodox welfare economics by using preference satisfaction as the underlying concept of advantage. To fit with the conception of the good he recommends—in terms of rational wants and his account of 'social primary goods'—the relevant preferences must be 'rational'. Otherwise, preference satisfaction would not necessarily be advantageous. And as with Harsanyi's view, there may be a divergence between Pareto efficiency in terms of actual and rational preference satisfaction. That opens up the possibility once more that QME might recommend 'paternalistic' interventions when people fall short in terms of rationality. These interventions would, once more, be justified with a view to correcting 'behavioural market failures'.

What does Rawls say about market failure? He moves on from his exposition of the first welfare theorem to the following remarks:

> The theory of general equilibrium explains how, given appropriate conditions, the information supplied by prices leads economic agents to act in ways that sum up to achieve this outcome [where no further mutually advantageous trades are possible]. Perfect competition is a perfect procedure with respect to efficiency. Of course, the requisite conditions are highly special ones and are seldom if ever fully satisfied in the real world. Moreover, market failures and imperfections are often serious, and compensating adjustments must be made by the allocation branch. (Rawls 1972, p. 272).

In this context, Rawls gives various 'branches' of government distinct jobs. Of these, the role of addressing market failure falls to the 'allocation branch' (Rawls 1972, p. 244). The 'allocation branch' is 'to keep the price system workably competitive and to prevent the formation of unreasonable market power'. But it is also 'charged with identifying and correcting, say by suitable taxes and subsidies and by changes in the definition of property rights, the more obvious departures from efficiency caused by the failure of prices to measure accurately social benefits and costs' (Rawls 1972, p. 244). The kinds of market failure he has in mind include:

'monopolistic restrictions, lack of information, external economies and diseconomies and the like' including public goods (Rawls 1972, p. 272). This list follows traditional welfare economics.

Ought the list of interventions which the 'allocation branch' should pursue be extended in the light of the results of BE to include libertarian paternalist interference? While we can only speculate about how Rawls might have responded to this question, two aspects of his view are particularly relevant here: the conditions under which Rawls thinks that 'paternalistic' principles might be accepted in the original position; and the definition of 'rationality' he adopts. Rawls (1972, pp. 248–250) discusses the 'problem of paternalism' and, to clarify what he has in mind, in a note he directs the reader to Gerald Dworkin's discussion, so that the definition of 'paternalism' he implicitly adopts is 'hard'. He notes that '[i]n the original position the parties assume that in society they are rational and able to manage their own affairs' (Rawls 1972, p. 248). He is concerned with the possibility that this assumption might not hold. Cases where it fails include those of children and those who are 'mentally disturbed'. As regards the parties in the original position, Rawls adds:

> It is rational for them to protect themselves against their own irrational inclinations by consenting to a scheme of penalties that may give them a sufficient motive to avoid foolish actions and by accepting certain impositions designed to undo the unfortunate consequences of their imprudent behaviour. For these cases, the parties adopt principles stipulating when others are authorised to act in their behalf and to override their present wishes if necessary; and this they do by recognising that sometimes their capacity to act rationally for their good may fail, or be lacking altogether (Rawls 1972, p. 249).

On the basis of whose preferences or conception of advantage would third parties act on their behalf? Rawls' answer is:

> Paternalistic decisions are to be guided by the individual's own settled preferences and interests insofar as they are not irrational, or failing a knowledge of these, by the theory of primary goods. As we know less and less about a person, we act for him as we would act for ourselves from the standpoint of the original position. We must be able to argue that with the development or recovery of his rational powers the individual in question will accept our decision on his behalf and agree with us that we did the best thing for him. (Rawls 1972, p. 249).[12]

This passage arguably supports the case for libertarian paternalist interventions within Rawls' account.[13] Those interventions are, like the 'paternalistic principles' Rawls has in mind, motivated and justified by failures of rationality, and (consistent with Rawls' discussion) the criterion used for judging a person's advantage is

[12] Rawls (1972, p. 250) qualifies these remarks by noting that: '[p]aternalistic principles are a protection against our own irrationality, and must not be interpreted to license assaults on one's convictions and character by any means so long as these offer the prospect of securing consent later on'.

[13] On this point, see also Ferey (2011, p. 747–748).

that person's own informed or rational preferences, as far as these are known. The FLoBIR and the 'paternalistic' interventions it implies would nonetheless only have a place within a contractarian account if there is some reason—which contractarian theory would have to supply—to believe that the parties to the contract would agree to it.

What standard of rationality is used to decide on whether or not 'paternalistic' intervention is justified in Rawls' theory? What conception of rationality does he adopt? The conception is, at least in part, that embodied in RCT. He writes: '[t]he concept of rationality invoked here, with the exception of one essential feature, is the standard one familiar in social theory. Thus, in the usual way, a rational person is thought to have a coherent set of preferences between options open to him. He ranks these options according to how well they further his purposes' (Rawls 1972, p. 143).[14] In the context of what makes a life plan 'rational', Rawls adds to the principles of rational choice that the plan should be chosen with 'full deliberative rationality, that is, with full awareness of the relevant facts and after careful consideration of the consequences' (Rawls 1972, p. 408). This second requirement is closer to the sort of requirement standardly adopted in informed preference views. Adding to this, the requirement that people have a 'coherent set of preferences' and that they act on these according to 'the principles of rational choice' nonetheless opens the way to 'paternalistic principles' if they fail to act on the axioms of RCT. Indeed, it opens the door to accepting the FLoBIR as a 'paternalistic principle' in the original position, or as a guiding principle to justify government intervention by the 'allocation branch' to address 'behavioural market failures'. Nonetheless, Rawls is not specific about the principles of rational choice which are involved. He merely cites a large literature on social and rational choice in a note (see Rawls 1972, p. 143). In his contractarian account, the standard of rationality would implicitly be set by the parties in the original position, since that standard would determine the principles they accept. The question is: would they set the standard of rationality at a level where they would accept the FLoBIR?

In thinking about this question, it is worth considering two further questions: how much knowledge of economics does Rawls deploy in developing his theory? And what knowledge do the parties in the original position have? Rawls is very modest about his knowledge of economics. He notes that '[c]ertain elementary parts of economic theory are brought in solely to illustrate the content of the principles of justice. If economic theory is used incorrectly or if the received doctrine is itself mistaken, I hope that for the theory of justice no harm is done' (Rawls 1972, p. 265). By contrast, the parties behind the 'veil of ignorance' have considerable knowledge:

> It is taken for granted, however, that they know the general facts about human society. They understand political affairs and the principles of economic theory; they know the basis of social organisation and the laws of human psychology. Indeed, the parties are presumed to know whatever general facts affect the

[14] The 'exception of one essential feature' here refers to the fact that Rawls (1972, p. 143) rules out envious preferences.

choice of the principles of justice. There are no limitations on general information, that is, on general laws and theories…' (Rawls 1972, p. 137).

Yet, in the context of Sugden's recent work where there is considerable controversy about which direction normative economics might take and about what we should learn from BE and 'the laws of human psychology', the parties may not be able to take any settled view on the basis of their knowledge.

Since there is no limitation on 'general information', the parties behind the veil and contractarians in the Rawlsian tradition would, nonetheless, look for answers to studies in economics and psychology which investigate how people view libertarian paternalist interventions. If they do, they will find that Sunstein (2016, p. 157) suggests that 'there is widespread cross-national support for nudges, at least of the kind we find that democratic societies have adopted or seriously considered in the recent past'. Nonetheless, Arad and Rubinstein (2018, p. 331) find that as regards 'the public's attitude toward the intervention methods advocated by libertarian paternalism … the responses provide several indications of a negative attitude'. Because contractarian theories require agreement, the presence of a significant number of negative responses to libertarian paternalistic interventions suggest that when contractarians update their theories in the light of empirical research, they cannot safely assume that contracting parties would agree to the FLoBIR.

We can now return to Sugden's claim that a 'contractarian cannot be a paternalist'. Rawls's theory appears to offer a counter-example to this claim, since he is a contractarian who believes that certain 'paternalist principles' would be agreed by the parties. In expressing this view, Rawls implicitly adopts Dworkin's 'hard' definition of 'paternalism'. Sugden's position shares a certain amount with Rawls' because they are both contractarians. Indeed, Sugden (2018a, pp. 260–261; see also Sugden 1989, pp. 74–79) introduces Rawls' theory as an exemplar of a contractarian theory of morals.[15] Yet, once one locates his work in relation to Rawls' (original version of his) theory, Sugden would no doubt suggest that elements of Rawls' theory are based on an orthodox position in normative economics which needs to be re-examined in the light of the results of BE. Rawls follows traditional welfare economics in assuming that people have coherent preferences. Sugden (2018a, p. 5) begins his work by noting that the findings of BE undermine this assumption. As a consequence, in The Community of Advantage Sugden reconsiders and (in chapter 6) modifies the case for the market in normative economics, and also (in chapter 7) develops his own view of regulation. Sugden also favours an opportunity-based over an informed (or 'considered') preference view of advantage (in chapter 5; see also Sugden 2006a), not least because he believes that traditional normative economics has assumed stable preferences—another assumption which has been challenged by BE (Sugden 2018a, p. 5) but also, no doubt, because he raises worries about some informed desire views (see Sugden 2000). Indeed, while in The Community of Advantage Sugden (2018a, p. 4) chooses J.S. Mill as a 'spokesperson' for the liberal tradition, he distances himself from those of Mill's views about well being which

[15] Sugden's discussions of Rawls include (Sugden 1989, 2010b) inter alia.

are associated with the informed desire view. He thinks Mill's views are those of a 'high-minded and intellectual humanist' and he thinks that some of these 'seem to rest on shaky foundations' (Sugden 2018a, p. 3).

In Sugden's account, it is 'in each person's interest to have opportunity to satisfy not just those preferences that she currently has, but any preferences she might come to have' (Sugden 2018a, p. 99). These preferences are not necessarily rational or informed and, given his anti-'paternalism', the government is not licenced to override people's preferences to address a failure of judgement by the individual with a view to furthering their good (Sugden 2018a, p. 46). On his account, some interventions advanced by LP are certainly (objectionably) 'paternalistic' (see Sugden 2008a, p. 230, 2017, 2018a, p. 46, 2018b; Sunstein 2018) and would be out of bounds in his theory, in part because they are addressed to a planner. On Sugden's contractarian view, LP 'lacks a valid addressee' because it is not addressed to 'individuals as directors of their lives' (Sugden 2018a, pp. 48–49). The structure and contents of the central parts of The Community of Advantage are organised with a view to offering a well-articulated alternative to LP and related 'paternalistic' proposals. For those contractarians who are considering whether 'paternalistic principles' might find a place in their account, Sugden (2018a, chapter 4, 2008a) offers a critique of 'paternalistic' proposals justified on the basis of the results of BE. By contrast, there is nothing in Rawls' approach as it is spelled out in A Theory of Justice which in itself blocks 'paternalistic' interference of this sort. Indeed, Sugden's claim that a contractarian cannot be a 'paternalist' does not generalise to other contractarian theories, such as Rawls'. It is specific to the sort of contractarianism that Sugden advances.[16] Indeed, for anti- 'paternalists' this point may count in favour of Sugden's view over Rawls'.

Clearly, Sugden's position differs from Rawls' in significant ways. It is important to note some of these differences because Sugden (2018a, pp. 261–262) suggests that 'what he has in mind is less grand in scope than Rawls' theory, but similar in spirit'. Nonetheless, in the light of the discussion here, Sugden would no doubt suggest that in the light of the findings of BE, Rawls' view of advantage should be amended; and to the degree that Rawls introduces the traditional framework of welfare economics into his discussion of the basic structure of society and of market efficiency and failure, he imports elements of, what Sugden thinks of as, a 'view from nowhere', and that the defence of the market and the implied view of regulation should be consistent with a contractarian view (as his own defence is) rather than based on traditional welfare economics.

While amendments of this sort may address some of the issues which Sugden might raise, Sugden's views and Rawls' theoretical framework may nonetheless diverge at a deeper level. Even if the view of advantage is amended, and the defence of the market and the view of regulation is revised, as we have seen, in Rawls' theory, 'paternalist principles' are agreed in the original position. This claim stands quite aside from Rawls' views about the currency of advantage, the market and regulation. For this reason, one must also ask whether the original position

[16] I am grateful to Robert Sugden for responses in correspondence on this topic on 22 and 24 May 2020.

device—which is central in Rawls' theory—is compatible with Sugden's views. In particular, is the view taken by the parties in the original position a 'view from nowhere'?

Rawls (1972, pp. 184–192) himself explains at length that the original position device is not used to model a 'rational and impartial sympathetic spectator' of the sort he associates with utilitarian views. As he puts it: '[f]rom the point of view of justice as fairness there is no reason why the persons in the original position would agree to the approvals of an impartial sympathetic spectator as the standard of justice' (Rawls 1972, p. 188). To the degree that Sugden (2018a, p. 19) has an 'impartial benevolent spectator' in mind when he talks of the 'view from nowhere', Rawls' position is not a 'view from nowhere'. Nonetheless, Sugden (2018a, p. 19) says of the 'view from nowhere' that its purpose is to 'try to filter out one's private interests and biases'. And this remark echoes Thomas Nagel's original discussion of (more or less) objective views in his book The View From Nowhere. There Nagel (1986, p. 5) writes that '[a] view or form of thought is more objective than another if it relies less on the specifics of the individual's make up and position in the world…'. As we saw, the parties behind the veil of ignorance do not know their place in society, in the distribution of natural abilities, their level of intelligence or strength, or their conception of the good and so on. To this degree, their view may be, following Nagel, more 'objective' than it otherwise would be. By contrast, Sugden (2006a, p. 209) insists that a 'contractarian understanding' must look for proposals which 'each individual can value from his or her point of view', and to this degree, it must 'treat social value as subjective'. Rawls' and Sugden's positions may again differ here. Furthermore, in Rawls' theory, 'instead of defining impartiality from the standpoint of the sympathetic observer … we define impartiality from the point of view of the litigants themselves … who must choose their conception of justice … in an original position of equality' (Rawls 1972, p. 190). Yet Rawls' goal in constructing the original position and the veil of ignorance is quite different to Harsanyi's aim in modelling the viewpoint of the impartial spectator. Rawls' purpose is to 'represent equality between human beings as moral persons' (Rawls 1972, p. 19). In Rawls' theory, as we saw earlier, the original position device is used to characterise the state of nature which is taken to be the status quo. Sugden's view may again differ significantly from Rawls' on this point. Sugden (2018a, p. 38) follows Buchanan in thinking that 'for contractarian thinking to be possible, it is sufficient that individuals acknowledge the baseline [or status quo] as a fact of life—that, as Buchanan puts it—"we start from here, and not from some place else" (1975, p. 78)'. Buchanan expresses himself in similar terms in his review of Rawls' A Theory of Justice. He writes that: '[w]e start always from here, not from an "original position"' (Buchanan 1972, p. 127). And Sugden (2018a, p. 174) concurs to the degree that—in contrast to Rawlsian 'justice as fairness'—he is clear that on his account 'for contractarian reasoning to be possible, it was not essential that the baseline was acknowledged as fair'. Indeed, on this line of reasoning, Rawls' early statement of his view is either not a contractarian view of the sort Sugden wishes to endorse, or at best a hybrid of

contractarianism and a 'view from nowhere'.[17] And yet, Rawls' theory is, arguably, the most influential contractarian theory in modern times. Perhaps for this reason, Sugden (2018a, p. 174) also notes that he deviates from 'the main paths of the contractarian literature, but in a direction previously taken by James Buchanan'. To this degree, there is a very significant difference between Rawls and Sugden, and Sugden may need to qualify his remark that his theory and Rawls' are 'similar in spirit'.

Even if there is a significant difference between Sugden and the early Rawls, there are clearly also actual or potential points of overlap between their views. The claim that the two theories are 'similar in spirit' rests on those points. As a contractarian, Sugden's view is consistent with Rawls' to the degree that 'in justice as fairness society is interpreted as a cooperative venture for mutual advantage' (Rawls 1972, p. 84). And while I have focussed on the original statement of Rawls' theory, it can be argued that there is more potential convergence between Sugden's view and Rawls' later view as it is set out in Political Liberalism and elsewhere.[18] In that view, Rawls takes 'the fact of reasonable pluralism' as given in a democratic society in which people hold distinct ('comprehensive') moral doctrines (see Rawls 1993, p. 135). Rawls' concern in 'political liberalism' is with a 'political conception of justice'—a moral conception worked out for the basic structure of society, which is taken to be a modern constitutional democracy (Rawls 1993, p. 11). In particular, he is concerned with a political conception which people who hold distinct doctrines can endorse, so that it is the object of an 'overlapping consensus' in a democratic society. In this context, Rawls argues that the principles of justice advanced in his theory are the objects of such a consensus. Sugden (2018a, p. 261) notes that his position would be close to this view if 'Rawls is representing his principles of justice as ones that actual people in an actual constitutional democratic society can agree to uphold'.

The idea of 'psychological stability' is also central in all versions of Rawls' theory[19] and Sugden (2018a, chapter 8) devotes an entire chapter to this topic in The Community of Advantage. Rawls (1972, p. 177) is concerned with a conception of justice which 'is stable when the public recognition of its realization by the social system tends to bring about the corresponding sense of justice'. In explaining Rawls' idea of 'psychological stability', Sugden (2018a, p. 174) writes that principles 'must be consistent with the facts of human psychology'. He adds that '[w]hen an ongoing society is regulated by those principles, it must reproduce both a general belief that the principles are fair and a general willingness to abide by them. Principles which are self-reproducing in this sense are psychologically stable'. But in The Community of Advantage Sugden (2018a, p. 174) is primarily concerned with the 'properties a

[17] This point also emerges implicitly in The Community of Advantage when Sugden contrasts the idea of a 'veil of uncertainty' (developed by Buchanan and Gordon Tullock) with the forms of 'veil' (which Harsanyi and Rawls) used to model impartial judgements. Sugden (2018a, p. 40) notes that '[t]he veil of uncertainty … is not a device for creating a view from nowhere'. See also Sugden (2018a, p. 284, note 4). I thank Robert Sugden for helpful responses on this topic in correspondence on 26 May 2020.

[18] For a comprehensive and original discussion of the relationship between Rawls' 'political liberalism' and Sugden's work, see Santori and Nalli (2019).

[19] In the case of his later views, see Rawls (1993, p. 17) and Rawls (2001, part 4) *inter alia*.

market economy needs to have in order for its governing principles to be psychologically stable viewed in a contractarian perspective'.

This convergence of Rawls' and Sugden's views also emerges in Rawls' related argument about the 'strains of commitment'. Rawls (1972, p. 176) notes that the parties in the original position 'cannot enter into agreements that may have consequences they cannot accept. They will avoid those they can adhere to with great difficulty'. Sugden (2018a, p. 194) would agree with Rawls that agreements must be such that 'we must be able to honor' them even if 'the worst possibilities prove to be the case'. But the central implication that Sugden draws from this line of argument, in the context of The Community of Advantage is that 'the contractarian recommendation in favour of a market economy needs to show each participant, looking ahead from where she is now, that she can expect the institutions of the market to work for her benefit. It needs to be able to do this, not just at some specially tailored starting line, but whenever "now" happens to be'. As he puts it, 'contractarian recommendations must engage with each individual's interests as she perceives them … If contractarian principles are to be psychologically stable that must mean each individual's interests as she currently perceives them…' (Sugden 2018a, p. 195). Here, again Sugden's position is consistent with Buchanan's view that we start from 'here' and not in some 'original position'. To this degree, while Sugden would acknowledge the convergence of his views with Rawls' (and also Rawls' influence on his views) on certain points, he is also very clear about where his position may differ from Rawls', especially as it is articulated in A Theory of Justice.

5 Conclusions

In The Community of Advantage Robert Sugden rejects the various forms of 'paternalism'—including LP—which have emerged by extending traditional welfare economics in the light of the results of BE. In the particular case of LP, that tradition has been extended with the use of a version of the informed (or rational) preference view of welfare. I have argued that while the logic of 'paternalistic' intervention in the context of 'behavioural market failures' or irrational preferences is prefigured in the views of J.S. Mill and John Harsanyi, there are convincing reasons why informed preference consequentialists should reject the agenda of LP, especially if like Harsanyi they endorse a form of rule-consequentialism. John Rawls' original articulation of his contractarian theory of justice adopts a version of the informed preference view. Because contracting parties in his theory agree certain 'paternalistic principles' to protect themselves from their own potential irrationality, Rawls' account appears to be a counter-example to Sugden's claim that contractarians cannot be 'paternalists'. Sugden's claim must be understood to be restricted in scope to the sort of contractarian view he himself advances. Indeed, 'anti-paternalist' contractarians may well, for this reason, favour Sugden's position to Rawls'. Rawls' discussions of rationality, market efficiency and failure in A Theory of Justice are based on orthodox welfare economics and may need to be updated in the light of the findings of BE. While Rawls' views, notably on psychological stability and the strains of commitment converge with Sugden's and shape some of the argument of

The Community of Advantage, to the degree that it can be argued that parties in the original position take a 'view from nowhere', Rawls' and Sugden's contractarian views are quite different. While contractarians in the Rawlsian tradition might be attracted by LP, Sugden's position offers a well-articulated 'anti-paternalist' contractarian alternative. Nonetheless, since 'anti-paternalist' consequentialists also have good reasons to reject the agenda of LP, Sugden's is not the only alternative to LP which liberals might explore.

Acknowledgements An earlier version of some parts of this paper was presented at a keynote talk at a conference on The Community of Advantage at LUMSA in Rome in November 2019. I thank Shaun Hargreaves-Heap, Robert Sugden and other participants at this event, and also an anonymous referee for their comments on earlier parts or versions of this paper. I also thank Federica Nalli and Paolo Santori for sharing their work with me. Most of all I thank Robert Sugden for his patient and helpful responses to my queries about his views. Any error is mine.

Compliance with ethical standards

Conflict of interest The author declares that he has no conflict of interest.

References

Arad A, Rubinstein A (2018) The people's perspective on libertarian-paternalistic policies. J Law Econ 61:311–333

Arrow K (1983) Collected papers of Kenneth J. Arrow, general equilibrium, vol 2. Basil Blackwell, Oxford

Bovens L (2008) The Ethics of Nudge. In: Grüne-Yanoff T, Hansson S (eds) Preference change: approaches from philosophy, economics and psychology. Springer, Berlin

Buchanan J (1972) Rawls on justice as fairness. Public Choice 13:123–128

Buchanan J (1975) The limits of liberty. University of Chicago Press, Chicago

Buchanan J (1986) Liberty, market and the state. Wheatsheaf, Brighton

Camerer C, Issacharoff S, Lowenstein W, O'Donohue S, Rabin M (2003) Regulation for conservatives: behavioral economics and the case for asymmetric paternalism. Univ Pa Rev 151:1211–1254

Conly S (2013) Against autonomy: justifying coercive paternalism. Cambridge University Press, Cambridge

Crisp R (1997) Mill on utilitarianism. Routledge, London

Dworkin G (1971) Paternalism. In: Wasserstrom RA (ed) Morality and the law. Wadworth, Belmont

Dworkin G (2020) Paternalism. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/spr2020/entries/paternalism/

Ferey S (2011) Paternalisme libéral et la pluralité du moi. Rev Econ 62:737–750

Glaeser E (2006) Paternalism and psychology. Univ Chic Law Rev 73:133–156

Griffin J (1986) Well-being: its meaning, measurement and moral importance. Oxford University Press, Oxford

Griffin J (2008) On human rights. Oxford University Press, Oxford
Harsanyi J (1981) Morality and the theory of rational behavior. In: Sen A, Williams B (eds) Utilitarianism and beyond. Cambridge University Press, Cambridge
Harsanyi J (1995) A theory of prudential values and a rule utilitarian theory of morality. Soc Choice Welf 12:319–333
Harsanyi J (1997) Utilities, preferences and substantive goods. Soc Choice Welf 14:129–145
Hausman D (2012) Preference, value, choice and welfare. Cambridge University Press, Cambridge
Hausman D (2018) Behavioural economics and paternalism. Econ Philos 34:53–66
Hausman D, Welch B (2010) To nudge or not to nudge. J Political Philos 18:123–146
Hooker B (2000) Ideal code, real world. Oxford University Press, Oxford
Le Grand J, New B (2015) Government paternalism: helpful friend or nanny state?. Princeton University Press, Princeton, New Jersey
Mas-Colell A, Whinston M, Green J (1995) Microeconomic theory. Oxford University Press, New York and London
Mill JS (1962) Utilitarianism. Including Mill's On liberty and essay on Bentham and selections from the writings of Jeremy Bentham and John Austin. Warnock M (ed). William Collins Sons & Co. Ltd., Glasgow
Nagel T (1986) The view from nowhere. Oxford University Press, Oxford
Osmani S (2018) Rationality, behavioural economics and Amartya Sen. J Hum Dev Capab 20:162–180
Qizilbash M (1998) The concept of well-being. Econ Philos 14:51–73
Qizilbash M (2006) Well-being, adaptation and human limitations. Philosophy 59:83–109
Qizilbash M (2011a) Sugden's critique of the capability approach. Utilitas 23:25–51
Qizilbash M (2011b) Sugden's critique of the capability approach and the dangers of libertarian paternalism. Int Rev Econ 58:21–42
Qizilbash M (2012) Informed desire and the ambitions of libertarian paternalism. Soc Choice Welf 38:647–658
Qizilbash M (2018) The market, utilitarianism and the corruption argument. Int Rev Econ 66:37–55
Rawls J (1972) A theory of justice. Oxford University Press, Oxford
Rawls J (1993) Political liberalism. Columbia University Press, New York
Rawls J (2001) Justice as fairness: a restatement. Kelly E (ed). Belknap Press of Harvard University Press, Cambridge, Mass. and London, England
Santori P, Nalli F (2019) The economic principle of political liberalism: a comparison of rawls and sugden. Paper presented at the 23rd annual conference of the European society for history of economic thought, Lille
Scheffler S (1988) Introduction. In: Scheffler S (ed) Consequentialism and its critics. Oxford University Press, Oxford
Sen A (1979) Utilitarianism and welfarism. J Philos 76:463–489
Sen A (1993) Markets and freedoms: achievements and limitations of the market mechanism in promoting individual freedoms. Oxf Econ Pap 45:519–541
Sen A (2000) Consequential evaluation and practical reason. J Philos 97:477–502
Sen A (2002) Rationality and freedom. Oxford University Press, The Belknap Press of Harvard University Press, Cambridge Mass
Sen A (2006) Reason, freedom and well-being. Utilitas 18:80–96
Sidgwick H (1881) The methods of ethics. Indianapolis and Cambridge, Hackett
Sugden R (1978) Social choice and individual liberty. In: Artis M, Nobay A (eds) Contemporary economic analysis. Croome Helm, London
Sugden R (1981) The political economy of public choice. Martin Robertson, Oxford
Sugden R (1985) Liberty, preference and choice. Econ Philos 1:213–229
Sugden R (1989) Maximizing social welfare: is it the government's business? In: Hamlin A, Pettit P (eds) The good polity: normative analysis of the state. Basil Blackwell, Oxford
Sugden R (2000) Review of Well-being and morality. Essays in honour of James Griffin Times higher education supplement: Friday 6 Oct
Sugden R (2004) The opportunity criterion: consumer sovereignty without the assumption of coherent preferences. Am Econ Rev 94:1014–1033
Sugden R (2006a) Taking unconsidered preferences seriously. Philo (R Inst Philos Suppl) 59:209–232
Sugden R (2006b) What we desire, what we have reason to desire, whatever we might desire: Mill and Sen on the value of opportunity. Utilitas 18:33–51
Sugden R (2008a) Why incoherent preferences do not justify paternalism. Const Polit Econ 19:226–248

Sugden R (2008b) Capabilities, happiness and opportunity. In: Bruni L, Comim F, Pugno M (eds) Capabilities and happiness. Oxford University Press, Oxford
Sugden R (2010a) Opportunity as Mutual Advantage. Econ Philos 26:47–68
Sugden R (2010b) Harsanyi, Rawls and the search for a common currency of advantage. In: Fleurbaey M, Salles M, Weymark J (eds) Justice, political liberalism and utilitarianism. Themes from Harsanyi and Rawls. Cambridge University Press, Cambridge
Sugden R (2017) Do people really want to be nudged towards healthy lifestyles? Int Rev Econ 64:113–123
Sugden R (2018a) The community of advantage. a behavioural economist's defence of the market. Oxford University Press, Oxford
Sugden R (2018b) Better off, as judged by themselves: a reply to Cass Sunstein. Int Rev Econ 65:9–13
Sugden R (2020) Normative Economics without preferences. Int Rev Econ. https://doi.org/10.1007/s12232-020-00356-8
Sunstein C (1991) Preferences and politics. Philos Public Aff 20:3–34
Sunstein C (2014) Why nudge? The politics of libertarian paternalism. Yale University Press, New Haven and London
Sunstein C (2016) The ethics of influence: government intervention in the age of behavioral science. Cambridge University Press, Cambridge
Sunstein C (2018) Better off, as judged by themselves: a comment on evaluating nudges. Int Rev Econ 65:1–8
Sunstein C, Thaler R (2003) Libertarian paternalism is not an oxymoron. Univ Chic Law Rev 70:1159–1202
Sunstein C, Thaler R (2005) Libertarian paternalism. In: Sunstein C (ed) The laws of fear: beyond the precautionary principle. Cambridge University Press, Cambridge
Thaler R, Sunstein C (2008) Nudge: improving decisions about health, wealth and happiness. Penguin Books, New York