What should economists do now?

Robert Sugden

School of Economics, University of East Anglia, Norwich NR4 7TJ, UK

r.sugden@uea.ac.uk

This essay is intended for inclusion in Richard E. Wagner (ed.), *James M. Buchanan: A Theorist of Political Economy and Social Philosophy* (in Palgrave Macmillan series: Remaking Economics: Eminent Post-War Economists)

Abstract This essay discusses the arguments in Buchanan's 1964 paper, 'What should economists do?', in the light of recent developments in behavioural economics. Criticizing the preference-satisfaction criterion of neoclassical economics, Buchanan argues that the central concern of economics should be to design and maintain institutions that allow individuals as much opportunity as possible to make their own choices and to engage in voluntary cooperation. He worries that using preference rather than choice as the fundamental normative concept might license social planners (and economists who see themselves as their advisers) to set themselves up as the judges of what individuals 'truly' prefer. This is exactly what behavioural welfare economics is now doing by using the satisfaction of (supposed) latent preferences as its criterion. Following Buchanan, I propose an opportunity-based criterion and argue that the contractarian justification for this criterion is unaffected by behavioural findings.

Keywords James Buchanan; contractarianism; opportunity; behavioural economics; paternalism.

Acknowledgements Some of the ideas presented in this essay were developed in discussions with Malte Dold, Alan Hamlin, Shaun Hargreaves Heap and Christian Schubert. My work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 670103.

Short biography Robert Sugden is Professor of Economics at the University of East Anglia, Norwich. His research uses a combination of theoretical, experimental and philosophical methods to investigate issues in behavioural economics, normative economics, the foundations of decision and game theory, the methodology of economics, and the evolution of conventions. His recent books include *Experimental Economics: Re-thinking the Rules* (Princeton University Press, 2009, with five co-authors) and *The Community of Advantage: A Behavioural Economist's Defence of the Market* (Oxford University Press, 2018). His current research focuses on the problem of reconstructing normative economics to make it compatible with behavioural findings.

What should economists do now?

‘What should economists do?’ is the title of the 1962 presidential address given by James Buchanan to the Southern Economic Association (Buchanan, 1964). It is a combative critique of what was then the prevailing understanding of economics – that economics is the science of constrained maximization, or of rational choice. Buchanan defends an alternative view of the discipline in which individual freedom and voluntary exchange are central concepts. In Buchanan’s framework, the concept of preference is redundant: there is no useful sense in which individuals have preferences, independently of their actual choices. Conventional welfare economics, in which the normative criterion is the satisfaction of preferences, is therefore fundamentally misguided. My essay revisits Buchanan’s arguments in the light of subsequent developments in behavioural economics, particularly the normative analysis that most behavioural economists have favoured and my own attempts to develop an alternative approach.

My work as an economist has been deeply influenced by Buchanan ever since I met him in 1977. I was then a very junior summer visitor to the Center for Study of Public Choice at Virginia Polytechnic Institute. Buchanan had invited me after reading and liking a paper I had written, criticizing Amartya Sen’s formalization of the concept of individual liberty.[1] During my stay at the Public Choice Center, I came to see how my critique of Sen’s theoretical framework fitted into Buchanan’s *contractarian* conception of normative economics. Ever since that visit, I have thought of my own work in normative and philosophical economics as contractarian in the sense that Buchanan expresses in ‘What should economists do?’

Since the early 1980s, I have also seen myself as working in what has come to be called ‘behavioural economics’. The distinguishing feature of behavioural economics is that, in explaining economic behaviour, it draws on ideas and research methods from cognitive psychology. I have never seen any tension between these two aspects of my work. To the contrary, I have come to see David Hume’s *Treatise of Human Nature* (1739-40/ 1978), which I first read on Buchanan’s recommendation, not only as one of the greatest works in the liberal tradition of social thought, but also as a founding contribution to cognitive psychology.[2] However, I have been increasingly disturbed by the tendency for behavioural

---

[1] The ideas in that paper appear in a more fully worked-out form in Sugden (1985).

[2] I defend this view of Hume’s *Treatise* in Sugden (2006).

economists to present their empirical findings as supporting, or even as necessitating, a paternalistic form of normative economics. I will say more about this tendency later .

There can be no doubt that behavioural research sets a serious challenge for neoclassical welfare economics. In neoclassical welfare economics, the normative criterion is the satisfaction of individuals' preferences. Each individual's preferences are assumed to be reasonably stable over time and to be independent of 'irrelevant' contextual features of the decision problems in which they are revealed. By virtue of these assumed properties, preferences are taken to be indicators of the individual's settled judgements about the value to her of alternative consumption options, or about the contributions those options would make to her well-being. But a large body of evidence shows that individuals' economic decisions often vary according to contextual cues that have no plausible relevance for welfare, but whose influence on decision-making can be explained by well-established psychological principles.

This evidence makes it difficult to defend (or even to use) preference-satisfaction as a normative criterion. But does it force economics to be paternalistic? I have spent much of the last fifteen years developing a form of normative economics that is compatible with the findings of behavioural economics, but is contractarian rather than paternalistic. Crucially, preference-satisfaction is not used as a normative criterion; instead, the criterion is the availability of opportunities for voluntary transactions.[3] I have claimed that my approach is broadly in continuity with Buchanan's. In this paper, I try to flesh out this claim by going back to Buchanan's texts.

1. Buchanan's vision of economic order

In 'What should economists do?', Buchanan's chosen adversary is Lionel Robbins. Buchanan takes issue with Robbins's famous definition of the 'economic problem' – the central subject-matter of economics – as the allocation of scarce means among alternative or competing ends. Buchanan points out that this definition says nothing about *whose* ends are relevant for economics. The implication is that economics is about constrained maximization or rational choice, considered in general. Buchanan's response is that concern with allocation problems *per se* 'is not a legitimate activity for practitioners of economics, as I want to define the discipline'; it is 'applied mathematics' or 'managerial science', not economics (1964, pp. 32–33).

[3] This work is brought together in Sugden (2018).

As I read him, Buchanan is not saying that models of optimization should have no place in economics. On any plausible account of what we economists should do, our job description includes trying to understand the workings of resource-allocation systems in general, and of the price system in particular. Buchanan has certainly seen his own work in this way. He has described how, as a graduate student at the University of Chicago, he was 'converted into strong advocacy of the market' by Frank Knight's price theory lectures, and how his work at the University of Virginia between 1957 and 1968 focused on promoting 'understanding of the price system' (1986a, pp. 3, 10). There are many areas of economic life where the immediate problems faced by individuals can usefully be modelled as the maximization of given objective functions subject to known constraints. It would be unreasonable to deny that neoclassical models of rational choice can throw light on significant features of human behaviour in markets.

Buchanan seems to acknowledge this in his discussion of the model of perfect competition. He says that this model has a 'basic flaw' – its 'conversion of individual choice behaviour from a social-institutional context to a physical-computational one' (1964, p. 36). I think that what Buchanan has in mind is that, in the model of perfect competition, there is no direct interaction between economic agents. Each individual's decision problem can be represented as that of maximizing a utility function subject to constraints imposed by technology and by market prices. The decision problems of different individuals are connected only through prices which, although endogenous to the model as a whole, are taken as given by each individual. As a modelling device, this representation enormously simplifies the analysis of a complex network of economic interactions. Buchanan is perhaps recognizing the legitimacy of this kind of modelling when he says that what is wrong with the model of perfect competition 'is not its lack of correspondence with observed reality; no model of predictive value exhibits this' (1964, p. 36). His objection is that the model provides a misleading framework for thinking about how markets work and what they do. It encourages the thought that a competitive market *just is* a solution to a set of equations – that we can understand the market without considering the actual process by which equilibrium is reached.[4] For Buchanan (as for Friedrich Hayek [1948], who makes a similar argument as part of his analysis of the limitations of central planning), that process is fundamental:

[4] A similar thought is implicit in David Gauthier's (1986, pp. 83–112) account of the market as a 'morally free zone'. Gauthier's contractarian theory treats a competitive market as morally equivalent to an archipelago of mutually isolated one-person island economies: 'Each person is thus a Robinson Crusoe, even in the market' (p. 91).

> A market is not competitive by assumption or construction. … [An equilibrium] solution, if there is one, *emerges* as a result of a whole network of evolving exchanges, bargains, trade, side payments, agreements, contracts which finally at some point, ceases to renew itself. At each stage in this evolution towards solution, there are *gains* to be made, there are exchanges possible, and this being true, the direction of movement is modified. (1964, pp.36–37)

This is a picture of the market as a process in which individuals are free to engage in mutually beneficial interactions. Competitive equilibrium is understood as a state in which all opportunities for such interactions have been exhausted.

Buchanan's fundamental opposition to Robbins is not about the kinds of models that we economists should or should not use. In this sense, it is not really about what we should *do*. It is about how we should understand what we are doing. What, ultimately, is economics *about*?

Buchanan's answer is that economics is about exchange rather than choice. The 'idea that should be central to our discipline' is that of exchange as 'a unique sort of relationship, that which involves the cooperative association of individuals, one with another, even when individual interests are different' (1964, p. 35). Or, as he puts it in another paper, economics provides 'an understanding of the social process through which a society of free persons can be organized without overt conflict while at the same time using resources with tolerable efficiency' (1986, p. 15). Notice how the second quotation suggests an alternative definition of the 'economic problem', as a problem of institutional or constitutional choice for the members of a free society.

There is a delicate balance here between normative and descriptive argument. Buchanan has a conception of an economy based on relations of voluntary cooperation between individuals who recognize both the separateness of their interests and the existence of opportunities for them to realise mutual benefit. He clearly believes that this is a desirable form of economic order, but (as I understand him) he sees the role of economics as to help us understand how, as a matter of empirical fact, such a system works. For my part, I would not go so far as to say that this is how *all* economists ought to understand their subject matter, but it is a large part of how *I* understand it.

2. Buchanan *versus* Robbins

If we accept Buchanan's idea that economics is about exchange, how exactly does Robbins's definition lead us astray?

Most obviously, economics goes astray if it treats the combined outcome of the decisions made in a society of interacting individuals as if it were the solution to a single maximization problem. This thought is the core of Buchanan's (1954) early and under-appreciated critique of Kenneth Arrow's Impossibility Theorem. Arrow's theorem is about the construction of a 'social' ranking of possible outcomes for society, using data about individuals' preferences over those outcomes. It is crucial to the impossibility result that the social ranking of outcomes is required to have 'collective rationality' properties that are formally similar to those that neoclassical economics attributes to individuals' preferences. Buchanan argues that there is no good justification for that requirement:

> Social rationality [for Arrow] appears to imply that the choice-making process produce results which are indicated to be 'rational' by the ordering relation, that is, the social welfare function. But why should this sort of rationality be expected? … Rationality or irrationality of the social group implies the imputation to the group of an organic existence apart from that of its individual components. (1954, p. 92)

My critique of Sen's analysis of individual liberty – the critique that won Buchanan's approval back in 1977 – followed a similar logic: Sen was trying to formulate the idea that society should respect individual liberty, using a conceptual framework that included requirements of collective rationality.

To require that 'social choice' is collectively rational is to treat the combined outcome of many individual choices as if it were the choice of a single agent. As Buchanan has often said (usually citing Knut Wicksell as the originator of this critique), economists characteristically address their recommendations to some imagined single agent who takes decisions on behalf of society – the 'social planner', the 'policy-maker', the 'government'. Since the recommendations themselves are supposed to maximize social welfare, the agent to whom they are addressed must be assumed to have both the desire and the power to do whatever is necessary to maximize welfare. In other words, the imagined addressee is a benevolent despot (Buchanan, 1986a, p. 23). If one's conception of society is of voluntary interaction between free individuals, the idea that it ought to work as if were under the control of a benevolent despot is deeply anomalous.

To use the model of the benevolent despot when making a normative appraisal of the market is to think of the market as a mechanism for solving an 'economic problem' in Robbins's sense – as the means by which some independently specified social objective might be achieved. For Buchanan, this way of thinking is fundamentally mistaken:

> The 'market' or market organization is not a *means* toward the accomplishment of anything. It is, instead, the institutional embodiment of the voluntary exchange processes that are entered into by individuals in their several capacities. That is all there is to it. (1964, p. 38).

But there is more to Buchanan's objection to Robbins's definition of economics than a rejection of the fiction of the social planner. The most basic normative principle in neoclassical welfare economics is the Pareto principle – that if, for some pair of social states *x* and *y*, every member of society weakly prefers *x* to *y* and at least one member strictly prefers *x*, then *x* is better for society than *y*. It is not self-evident that 'better for society' should be read as 'better, as judged by a social planner'. Might it not instead be read as 'better, as agreed by all members of society'? On this reading, one might think, the Pareto principle uses the idea that *individuals* have rational preferences without presupposing anything about *social* rationality. Extending this idea, one might think of the concept of individual preference as one of the basic building blocks of any liberal form of normative economics. This thought becomes particularly plausible if, moving beyond the kind of mathematics used by Robbins, one considers the role of (non-cooperative) game theory in present-day economists' models of human interaction. Game theory is based on maximizing principles of *individual* rationality that are stronger than the ordinal principles invoked by Robbins: the concept of 'payoff' used in game theory presupposes that individuals' preferences satisfy the axioms of expected utility theory. (Thus, utility is assumed to be cardinal, but there is no assumption of interpersonal comparability.) But the 'solution concepts' used in game theory are intended to represent the players' common knowledge of each other's individual rationality; there is no assumption that the solution to a game is *collectively* rational in Arrow's sense. Think of the Prisoner's Dilemma. Defection by both players is normally understood to be the uniquely correct solution to this game – the only combination of strategies that is consistent with common knowledge of rationality – even though it is clearly sub-optimal from a social point of view.

Nevertheless, 'What should economists do?' is written as an objection to the maximizing perspective *in general*, and not merely to the idea of maximizing a social welfare function. Buchanan is more explicit about this in a later work, in which he argues that economics took a wrong turn after Alfred Marshall. As in the 1964 paper, that wrong turn is characterized as the search for 'maximizing or optimizing solutions within the constraints of specific wants, resources and technology', which Buchanan now calls the 'mathematical perspective'. Explaining what is wrong with this perspective, he says:

> I suggest that the mathematical perspective takes hold once we so much as define persons as utility or preference functions and implicitly presume that these functions exist independently of the processes within which persons make actual choices… By postulating such functions independently, and by imposing the resource constraints, it then becomes possible to define, at least conceptually, the 'efficient' allocation of resources, quite apart from any voluntary process of agreement among trading parties. This formalization of the efficiency norm then allows the market to be conceptualized as merely a means, a mechanism, one among others, to be tested or evaluated in terms of its efficacy in attaining desired results in the utilization of resources. (1986, pp. 16–17)

What Buchanan is objecting to here is an assumption that he takes to be implicit in most forms of neoclassical economics – the assumption that an individual's preferences can be identified prior to decision problems in which they can be expected to be revealed.

An economist who takes the 'mathematical perspective' might question whether economics *does* assume this. She might point out that the economic concept of preference is understood in relation to choice. According to one common account of the relationship between preference and choice, a preference for x over y is a prevailing disposition to choose x rather than y. Alternatively, according to a strict revealed-reference account, to say that a person prefers x to y is to say that she would not choose y from an opportunity set that contained x. On either account, the economist's claim to knowledge about a person's preferences at any given time is based on evidence about what that person has chosen in the past, or on other information that is relevant for predicting how she will choose in the future. Nevertheless, Buchanan's claim is correct. Whenever economists use propositions about preference to explain or predict individuals' choices, they are endorsing a conceptual distinction between preference and choice – the distinction between what justifies one in making a prediction and what makes that prediction true or false.

If I am reading Buchanan rightly, he does not deny the potential usefulness of preference-based models as aids to predicting individual behaviour. But when what is at issue is the normative evaluation of institutions, he sees the separation of preference and choice as inconsistent with his favoured conception of individual liberty. This alleged inconsistency has two sources, which I will consider in turn.

3. Liberty as individual sovereignty

When Buchanan (1986a, 1986b) writes about his own intellectual development, he describes his political position before going to the University of Chicago in 1945 as 'libertarian socialist' or 'populist'. Explaining the libertarian component of this position, he says:

> The person who shares this perspective places a primary value on *liberty*, as such. He personally disputes, rejects, resents, opposes attempts by others to exercise control or power over his own choice behaviour. He does not like harness. There is an exhilaration in simply being free. (1986a, p. 4)

This attitude to liberty, he says, is encapsulated in the words 'Don't tread on me' on the rattlesnake banner flown in the American War of Independence (1986a, p. 5). Socialism and populism enter the picture through the young Buchanan's identification of those who might exercise power over him as 'robber barons' (people who, he says, seemed very real to him in 1945) and the 'Eastern establishment' (with which even the later Buchanan was never quite able to reconcile himself). The point of the story is that Buchanan's socialism was short-lived, but his commitment to libertarianism continued throughout his life.

'Don't tread on me' expresses a first-person conception of liberty. Liberty is something that I can *demand*. I do not need to give reasons why my having liberty is fitting for me as a human being, or why this is good for society as a whole, or even why it is good for me. No one has the standing to ask me for reasons. It is enough that I want liberty and am entitled to have it. In more philosophical vein, Buchanan describes this position as 'normative individualism':

> The justificatory foundation for a liberal social order lies… in the normative premise that individuals are the ultimate *sovereigns* in matters of social organization, that individuals are the beings who are entitled to choose the organizational-institutional structures under which they will live… If individuals are considered the ultimate sovereigns, it follows directly that they are the *addressees* of all proposals and arguments concerning constitutional-institutional issues' (1991, p. 288).

The sentence about addressees expresses what I take to be the fundamental principle of Buchanan's contractarianism – that normative analysis is addressed to individual citizens as potential parties to mutually beneficial agreements. The relationship of citizen to government is not that of passive subject to benevolent ruler; it is that of principal to agent.

If one thinks in terms of a principal–agent relationship, it is easy to understand why an individual citizen might be reluctant to allow public decisions to be justified in terms of claims about what he and his fellow-citizens prefer. For a citizen to allow this would be for him to transfer normative authority from himself as sovereign chooser to someone else (let us call her the 'social planner') who acts as the judge of what he prefers. It opens up the possibility that the planner might use her own judgements about the citizen's true preferences to overrule his actual choices. Buchanan characterizes the standard approach to normative economics as a construction in which there is an 'ontological assumption that there is "something" – whether called a utility function or not – that exists and can, at least conceptually, be objectified and separated from individual choice'. If this assumption is made:

> [The] relationship between an individual's choice behaviour and his or her utility function becomes a matter of fact. That is, there arises a factual question open to investigation concerning the correspondence between the choices made and the change in the individual's position as measured on the independent scalar… [And then it makes sense] to raise the question as to whether the individual or some third party or parties can most reliably identify the choices that are defined as 'best' in terms of the given utility function. (1991, p. 283)

Buchanan is committed to a different assumption: 'My own ontological presuppositions do not allow any conceptual separation or distinction between an individual's choice behaviour and his or her utility function… All there is are individual choices' (1991, p. 286).

Thus, for Buchanan, an individual's liberty should not be understood in terms of his getting what he prefers, with 'preferring' defined independently of 'getting'. His liberty should be understood in terms of *what he is free to choose*. This is a property of his choice set – the set of options between which he is able to choose. Crucially, an individual's choice set is defined without reference to his preferences. This line of thought leads Buchanan to the conclusion that it is in each person's interest that his choice set is 'as open as is naturally possible' (1979, pp. 258–259).

Taking the first-person view of liberty, you can demand that you are not prevented from choosing a certain option, say *x*, without having to give any reason for that demand, other than that *x* is something that you *might* want to choose. You can say that without actually having any current desire to have *x*. Thus, your demand cannot be countered by a social planner who judges, perhaps on the best possible evidence of your previous behaviour, that you have no current preference for *x* and are unlikely to have such a preference in the future. Notice that, in making this demand, you are asserting your sovereignty *now* to act on behalf of your future self, and the principle on which you are acting is that of trying to ensure that, in the future, you will be able to get what you *then* want – even if you do not yet know what that will be. You now want it to be the case that if, in the future, you want *x*, you will be able get x.[5] That seems to me to be an uncomplicated application of the principle of 'Don't tread on me'.

In some of his writing, however, Buchanan seems to want to connect the demand for freedom of choice to philosophically deeper notions of ontology and to morally deeper ideas about what it means to be human. I believe that these moves are not necessary for Buchanan's contractarian approach, and are liable to undermine it. Let me explain.

4. Liberty as autonomy

In 'What should economists do', Buchanan objects to the conception of human agency that he sees as built into the theory of rational choice:

> In one sense, the theory of choice presents a paradox. If the utility function of the choosing agent is fully defined in advance, choice becomes purely mechanical. No "decision", as such, is required; there is no weighing of alternatives. On the other hand, if the utility function is not wholly defined, choice becomes real, and decisions become unpredictable mental events. (1964, p. 34)

The idea seems to be that, because choice theory represents decisions as predictable (given data about preferences), it is modelling human agents as mere mechanisms, lacking in autonomy. The decisions of an autonomous agent, it is implied, would necessarily be unpredictable.

These thoughts are developed much further in Buchanan's 1979 paper 'Natural and artifactual man'. This paper has a more Austrian flavour than most of Buchanan's other work; he says that it was influenced by the work of George Shackle, which he had been reading at the time (Buchanan, 1979, p. 251). Buchanan wants to contrast 'natural man', understood as a being whose behaviour can be predicted and explained naturalistically, with 'artifactual man', understood as an actor whose decisions are autonomous and thereby unpredictable, and who views his future self as his own construction (1979, pp. 246–248).

As in his 1964 paper, Buchanan argues that rational choice theory represents individuals as non-autonomous:

> The rational ideal eliminates choice, as Shackle emphasizes. Choice requires the presence of uncertainty for its very meaning. But choice also implies a moral responsibility for action. To rationalize or to explain choices in terms of either

[5] This is a conception of individual agency as a *continuing locus of responsibility*. I say more about this in Sugden (2004).

> genetic endowment or social environment removes the elements of choice and responsibility. (1979, p. 257)

Notice that Buchanan is classifying rational choice theory together with other branches of natural and social science that claim to give empirical explanations of human behaviour. Such explanations, he claims, treat human beings as if *Homo sapiens* were just a 'natural animal', one animal species among many. But it is intrinsic to the nature of human beings that their behaviour cannot be fully predicted; and unpredictability is an essential part of moral responsibility:

> If individual man is to be free, he is to be held accountable, he is to be deemed responsible for his actions. But at the same time he is allowed to take credit for his achievement. Who can claim credit for results that could have been predicted from nature? From a knowledge of his genetic endowment or his social environment, or both? But once man is conceived in the image of an artifact, who constructs himself through his own choices, he sheds the animalistically determined path of existence laid out for him by the orthodox economists' model. (1979, pp. 257–258)

On this view, the whole idea of a science of human behaviour is morally objectionable.

The aspect of autonomy that is central to Buchanan's 1979 paper is the possibility of deliberately choosing the kind of person that you will become. He gives the example of spending on education, characterized as 'investing in becoming' – investing in creating 'the person that we want to be rather than the one we think we might be if the spending is not made in this way'. Musical appreciation is another example: through study and practice, you can invest in a way that will shift your preferences towards being appreciative of certain kinds of music (1979, pp. 248–249). Buchanan's approval of self-creation extends to forms of 'investment in becoming' by which you voluntarily subject yourself to constraints that you cannot then reverse, or authorize other people to coerce your future self against acting on its desires (p. 253).[6]

As long as trying to influence your future preferences is understood as just one of the many ways in which you might use freedom of choice, there is perhaps no tension between this line of thought and 'Don't tread on me'. But Buchanan takes a further step: he presents his picture of artifactual man as *the reason why liberty is valuable*.

---

[6] Approval of unilaterally self-imposed constraints is a recurring theme in Buchanan's work, aligned with his advocacy of constitutional rules to restrict day-to-day political decision-making. His analysis of the 'Samaritan's dilemma', for which the solution is a self-imposed rule against generosity to would-be 'parasites', is an example (Buchanan, 1975a).

> [Man wants liberty] precisely because he does not know what man he will want to become in time. Let us remove once and for all the instrumental defense of liberty, the only one that can possibly be derived directly from orthodox economic analysis. Man does not want liberty in order to maximize his utility, or that of the society of which he is a part. *He wants liberty to become the man he wants to become*. (1979, p. 259; italics in original)

And he tells us that we *ought* to want liberty for this reason: 'Individually, persons must recapture an ability to imagine themselves as "better" persons than they are' (1979, p. 254).

I think these are wrong moves for a contractarian who values liberty. A contractarian justification of liberty needs to show each and every individual that liberty is in his or her own interests, as he or she understands them. But if the value of liberty really is tied to the unpredictability of human choice, a person who believes that his own decisions are predictable will not be able to see why his own liberty is valuable to him. And if the true reason for wanting liberty is to act as an 'artifactual man', a person who does not think of his future self as his own artifact is left without a reason for wanting to be free. Remember that the addressees of a contractarian argument are individual citizens, viewed as sovereigns. Ultimately, the only reasons that matter are the reasons that those citizens actually accept. The greater the degree to which an argument depends on specific ontological claims or on specific ethical commitments, the less power that argument has. Viewed in this perspective, 'Don't tread on me' looks a more promising justification for liberty than self-creation.

Why might someone want liberty without subscribing either to the ontology or to the self-creation ethic of Buchanan's 1979 paper? With respect to the ontology, my answer is that predictability *is* compatible with autonomy. Autonomy, as I understand it, is a sense of volition. It is a person's subjective perception of herself as the cause of her own actions – her perception that she has the power to act in ways that in fact she chooses not to do. It is possible to have this sense about highly predictable decisions. Take a personal example. Over the last five years, I have faced hundreds of decision problems in which my choice set contained both coffee and Coke. In that time, I have chosen coffee hundreds of times and (I am fairly sure) never chosen Coke. This pattern of behaviour is consistent with the hypothesis that, with respect to those two drinks, I prefer coffee to Coke and make rational choices between them. The proximate explanation for this behaviour is that I like the taste of coffee and do not particularly like that of Coke (which I can remember from drinking it long ago). These tastes may well have physiological causes. But whatever the explanation, my choices between the two drinks feel autonomous to me. On any occasion on which I face this choice, I have no sense of compulsion to choose coffee: my internal sense is that I *could* choose Coke, I just don't want to. And because I can imagine wanting to choose Coke in future situations, I can want Coke to continue to be an option in my choice set. I can say to a social planner: Just because I haven't chosen Coke for five years, don't presume that I will never want to choose it in the future. Let me choose for myself. Don't tread on me.

What about the ethic of self-creation? I can accept that *some* people make conscious efforts to shape their future preferences, but liberty also matters to those who don't. As far as I can recall, I never thought of my own educational choices as attempts to change my preferences. When I went to university at the age of eighteen, I had a strong sense of what Buchanan describes as exhilaration in being free – free from the oversight of parents and schoolteachers, free to try out new experiences. I looked forward to enjoying new intellectual challenges and having new opportunities to exercise my abilities. And, like most of my contemporaries, I expected that getting a university degree would be a stepping stone to some fulfilling career from which I would earn a comfortable income. Although I had only hazy ideas about what exactly these opportunities would be, I was excited about what might lie ahead. All this gave me good enough reasons to value my liberty. If someone had said to me that I ought also to have been trying to improve my preferences, I could reasonably have replied that it was up to me to decide whether or not my preferences were in need of improvement.[7] Don't tread on me.

Treating unpredictability and self-creation as *the* reasons for valuing liberty reduces the constituency to which contractarian arguments for liberty can be addressed. It can also provide would-be social planners with arguments for restricting citizens' freedom of choice. From the idea that autonomy necessarily involves unpredictability, it is a short step to the idea that people whose choice behaviour is stable over time are not autonomous agents, but merely natural animals for whom liberty serves no purpose. From the idea that liberty has value precisely because it allows individuals to engage in self-creation, it is a short step to the idea that liberty has no value to individuals who make no effort to improve themselves. Dual-self models of self-constraint, in which a person's 'planning' self imposes constraints on his 'impulsive' self – the model that is implicit in the story of Odysseus and the Sirens – encourage would-be social planners to suppose that their fellow-citizens have latent desires

[7] In fairness to Buchanan's adversary, I should add that I was fortunate in being able to go one of the 'new universities' established in Britain in the 1960s on the recommendation of a committee chaired by the then Lord Robbins. My education there had all the properties that tend to foster self-creation.

for restrictions to be imposed on their choice sets. I do not want to claim that Buchanan's account of the connection between liberty, autonomy and self-creation directly justifies any constraints on choice other than those that each individual chooses to impose on herself. Still less do I want to claim that Buchanan favoured such constraints. But I do maintain that there is a tension between Buchanan's treatment of autonomy and his contractarianism.

It seems to me that the best way to resolve this tension is to distinguish between contractarian arguments and personal value judgements about a good society. When you write as a contractarian, your arguments are addressed to your fellow-citizens and are intended to engage with *their* judgements about *their* interests. But taking a contractarian perspective in your work as a normative economist does not debar you from expressing your own ideas about a good society – provided you acknowledge the distinction between the two activities. *In Buchanan's picture of a good society*, each person wants liberty to become the person he or she wants to become. And so, *in that imagined society*, each person wants its institutions to be structured so that each person has a rich range of opportunities for self-creation, including opportunities for Odysseus-like self-constraint. But unless actual people, here and now, want these things, all this just is a picture of an ideal world. There is a sense in which painting such pictures is what Buchanan (1975b, pp. 1–2) has called 'play[ing] at being God'. That is not an intellectual crime, but it is not contractarianism.

5. What should economists do with the findings of behavioural science?

Buchanan's arguments about what economists should do were written in a period when almost all economists were content to assume that individuals' decisions revealed well-behaved context-independent preferences. In denying the usefulness of the concept of preference, Buchanan was opposing a central feature of received economic theory. His critique of the way the concept of preference was used in normative economics was taken up by only a tiny minority of economists (of whom I was one). But since then, there has been a huge change in economists' understanding of decision-making behaviour. How far individuals' choices reveal well-articulated preferences is one of the central topics of behavioural economics, and behavioural economics is at least well on the way to becoming mainstream.

One of the fundamental findings (or rediscoveries[8]) of behavioural economics is that individuals' choices are often *context-dependent*. That is, an individual's choices between what economics has normally understood as 'given' options vary according to features of the decision environment that seem to have little relevance to the individual's interests or welfare, but whose effects are psychologically explicable. To give just a few examples: choices from given sets of options are systematically influenced by which option is described as the status quo and by manipulations that direct the chooser's attention to particular features of those options; preferences between given pairs of options differ according to whether they are elicited directly in choices or inferred from monetary evaluations elicited for the two options separately; preferences between 'smaller sooner' and 'larger later' payoffs reverse as the 'sooner' date approaches.[9]

On the most natural interpretation, this body of evidence confirms Buchanan's criticism of the 'mathematical perspective' of neoclassical economics – the criticism that it illegitimately assumes that preferences exist independently of the processes within which persons make actual choices. Ironically, however, the confirmation takes a form that Buchanan might not have welcomed. Recall that one of Buchanan's objections to the conventional theory of rational choice is that it treats human decision-making as predictable by the methods of empirical science; he thinks that this approach fails to recognize human choice as autonomous. Behavioural economics is much *more* empirical than rational choice theory. An advocate of rational-choice theory might claim that the reason why that theory can be expected to predict successfully is that human decision-makers really are autonomous rational agents: all that is being predicted about them is that their reasoning is consistent with the principles that define what is meant by 'rationality'. But if behavioural economics is grounded on empirical psychology (as its practitioners normally claim it is), it must use what Buchanan calls a model of 'natural man'. Given Buchanan's distrust of genetic and social-

[8] Many of these effects were known, to psychologists and to some economists, long before the explosion of interest in behavioural economics. (For example, Wicksteed's [1910] exposition of neoclassical economics includes psychologically acute discussions of many now-familiar 'anomalies'.) What is new is the widespread recognition of the economic significance of these effects.

[9] There is far too much evidence about these and other 'anomalies' for specific citations to be useful. A representative sample of this evidence is collected in Kahneman and Tversky (2000).

environmental explanations of choice behaviour, it is reasonable to guess that he would not have felt much sympathy for the programme of behavioural economics.[10]

For my purposes in this essay, what is particularly interesting about the way that behavioural economics has developed is its confirmation of one of Buchanan's criticisms of what his fellow-economists do. Recall his argument that if utility can be 'objectified and separated from individual choice', it becomes meaningful to ask 'whether the individual or some third party or parties can most reliably identify the choices that are defined as "best" in terms of the given utility function' (1991, p. 283). To an economist who is confident in the predictive power of rational-choice theory, Buchanan's concern might seem misplaced. If an individual's utility function really can be constructed from observations of her previous choies, and if that function really can predict her future choices, why should she worried if the identification of what is best for her is made by a 'third party' who knows what that function is? Of course, Buchanan's starting point was scepticism about the predictive power of rational-choice theory, given the autonomy of human decision-makers. But behavioural economics gives further reason for that scepticism, and therefore further reason for concern about allowing social planners to judge what is best for an individual.

Among behavioural economists, there is now a broad consensus about how normative analysis should be conducted. The essential idea can be traced back to Cass Sunstein and Richard Thaler's (2003) well-known story of the (hypothetical) cafeteria in which customers' choices between food items are influenced by the relative positioning of the items on the cafeteria counter. Other things being equal, more prominently displayed items are more likely to be chosen. The cafeteria director can choose what display to use. Treating this example as a microcosm of the problem of how to do welfare economics when individuals' choices are context-dependent, Sunstein and Thaler ask how the director should make this choice. They conclude that, because individuals lack well-defined preferences that the director could try to respect, the idea that there are 'viable alternatives to paternalism' is a 'misconception'. They argue that the director should be a *libertarian paternalist*: she should not constrain the customers' opportunities for choice, but she should display the items in the

[10] As far as I can recall, issues arising from behavioural economics did not feature in any of the many conferences at which Buchanan and I were co-participants. In his later years, I would have been reluctant to initiate discussion of topics on which I expected the two of us to have fundamental disagreements.

way that results in their making the choices 'that she thinks would make the customers best off, all things considered' (Sunstein and Thaler, 2004, pp. 1164–1165, 1182).

In their later book *Nudge*, Thaler and Sustein (2008) are more explicit about the criterion that normative economics should use. They say that their recommendations are designed to 'make choosers better off, *as judged by themselves*'. Expanding on this, they say that behavioural economics has shown that 'in many cases, individuals make pretty bad decisions – decisions that they would not have made if they had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control' (p. 5). The clear implication, confirmed over the subsequent course of the book, is that 'better off, as judged by themselves' is to be interpreted by reference to the preferences that the relevant individual would have revealed, had his decision-making not been impaired by limitations of attention, information, cognitive ability or self-control. In other words, normative analysis should try to reconstruct individuals' underlying or *latent* preferences by simulating what they would have chosen, had they not been subject to imperfections of reasoning and information. Notice the implicit assumption that latent preferences are not themselves context-dependent. (Remember that the point of using latent rather than revealed preferences is that revealed preferences are context-dependent.) As Gerardo Infante, Guilhem Lecouteux and I have argued, this method of analysis proceeds as if, inside each individual, there is an *inner rational agent* with neoclassical preferences; behavioural deviations from neoclassical theory are supposed to occur because of psychologically-induced 'errors' in the implementation of the agent's latent preferences (Infante et al., 2016).

In slightly different forms, this implicit model of an inner rational agent recurs in many attempts to derive normative conclusions from behavioural economics.[11] Some authors try to 'purify' revealed preferences by identifying and removing the alleged effects of errors of reasoning. Others try to identify the environments in which individuals' decisions are least prone to error and then infer individuals' preferences by using only data generated in those environments. Clearly, this strategy of *behavioural welfare economics* can work only if the concept of latent preference has an objective definition – if it is not just another word for the analyst's personal judgement about what is best for each individual – and only if latent preferences are context-independent. But the advocates of this strategy usually offer (at best) only rough sketches of how 'error' is to be defined operationally, and almost never try to

[11] This paragraph contains sweeping claims that some behavioural economists might dispute. The evidence on which they are based is presented in Infante et al., 2016 and Sugden, 2018).

justify the crucial assumption that, after the effects of error have been eliminated, preferences will be found to be context-independent. Viewed in the perspective of empirical psychology, the whole concept of latent preference seems question-begging and redundant. (If actual choices can be explained as the result of known mental processes, why should we expect there to be other mental processes, as yet undiscovered, that generate preferences that are *not* used in decision-making but which happen to have the properties that correspond with the axioms of rational-choice theory?) Thus, whatever its advocates may intend, the strategy of behavioural welfare economists leaves the social planner or her adviser with a large amount of discretion in specifying individuals' latent preferences, and very little useful guidance about how that discretion should be used.

I conclude that Buchanan was right to be worried about the conceptual distinction in neoclassical economics between preference and choice, and about the practice of using preference rather than choice as the fundamental normative criterion. His worry was that this practice might license social planners (and economists who think of themselves as advisers to social planners) to set themselves up as the judges of what individuals 'truly' prefer, and to design social institutions to satisfy those supposed preferences. That is exactly what behavioural welfare economics is now doing.

How else might normative economics be done? Buchanan offers an obvious answer – obvious, that is, to anyone who shares his commitment to individual sovereignty:

> [Each person] has a clear interest in seeing that the choice set, the set of alternative imagined futures, remains as open as is naturally possible, and, if constrained, that the constraints be also of his own choosing. (1979, pp. 258–259)

In other words, the normative criterion should be one of *opportunity*. Roughly speaking, the aim should be to set up institutions that give individuals as much opportunity as possible to do whatever they want to do, both in their actions as separate individuals and in voluntary transactions with one another. In designing these institutions, there is no need to consider what individuals' preferences in fact are: the aim should be to ensure that individuals are able to act on *whatever* preferences they may happen to have in any particular context, at any particular time. Or, as Buchanan might put it, there is no need to talk about preferences at all; all that matters are opportunities *to choose*. Whether or not an individual's choices can be rationalized by context-independent preferences is beside the point.

This general strategy can be followed in at least two different ways. One way is arguably in the spirit of Buchanan's arguments about autonomy and self-creation. Its aim is to design and maintain institutions that individuals *would* want, were they to attach sufficient importance to being free to engage in self-creation. Or to put this another way, its aim is to create a society in which individuality and self-creation can flourish. Some ideas about how this approach to normative economics might be developed can be found in the work of Shaun Hargreaves Heap (2013, 2017), Christian Schubert (2015) and Malte Dold (2018).

The other way forward is the one that I have been developing, and which is summarized in my book *The Community of Advantage* (Sugden, 2018). This way, I believe, is more faithful to the spirit of Buchanan's contractarianism and to the slogan of 'Don't tread on me'. It accepts that individuals' choices are often context-dependent in ways that psychological theories can predict and that neoclassical theory would classify as irrational, but does not interpret that fact as compromising individuals' autonomy as choosers or their claims to sovereignty. It views the market favourably, for the same reasons that Buchanan does when he says that the market is the institutional embodiment of voluntary exchange between individuals in their several capacities. The market, and civil society more generally, is not a means to the accomplishment of anything – not even a means to the flourishing of individuality. Voluntary interaction is, as Buchanan says, all there is to it.

References

Buchanan, James (1954). Social choice, democracy, and free markets. *Journal of Political Economy* 62: 114–123. Reprinted in Buchanan (1999).

Buchanan, James (1964). What should economists do? *Southern Economic Journal* 30: 213–222. Page references to reprint in Buchanan (1999).

Buchanan, James (1975a). The Samaritan's dilemma. In Edmund Phelps (ed.), *Altruism, Morality and Economic Theory*. New York: Russell Sage Foundation, 71–85. Reprinted in Buchanan (1999).

Buchanan, James (1975b). *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: University of Chicago Press.

Buchanan, James (1979). Natural and artifactual man. In James Buchanan, *What Should Economists Do?* Indianapolis: Liberty Fund, 93–112. Page references to reprint in Buchanan (1999).

Buchanan, James (1986a). *Liberty, Market, and State*. Brighton: Wheatsheaf.

Buchanan, James (1986b). Better than plowing. In *Banco Nazionale del Lavoro Quarterly Review* 159: 359–375. Page references to reprint in Buchanan (1999).

Buchanan, James (1991). The foundations for normative individualism. In James Buchanan, *The Economics and the Ethics of Constitutional Order*. Ann Arbor, MI: University of Michigan Press. Page references to reprint in Buchanan (1999).

Buchanan, James (1999). *The Collected Works of James M. Buchanan: Volume 1*. Indianapolis, IN: Liberty Fund.

Dold, Malte (2018). Back to Buchanan? Explorations of welfare and subjectivism in behavioral economics. *Journal of Economic Methodology* 25: 160-178.

Gauthier, David (1986). *Morals by Agreement*. Oxford: Oxford University Press.

Hargreaves Heap, Shaun (2013). What is the meaning of behavioural economics? *Cambridge Journal of Economics* 37: 985–1000.

Hargreaves Heap, Shaun (2017). Behavioural public policy: the constitutional approach. *Behavioural Public Policy* 2. Available online at: https://doi.org/10.1017/bpp.2016.8

Hayek, Friedrich von (1948). *Individualism and Economic Order*. Chicago: University of Chicago Press.

Hume, David (1739-40/ 1978). *A Treatise of Human Nature*. Oxford: Oxford University Press.

Infante, Gerardo, Guilhem Lecouteux and Robert Sugden (2016). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology* 23: 1–25.

Kahneman, Daniel and Amos Tversky (eds) (2000). *Choices, Values, and Frames*. Cambridge: Cambridge University Press and Russell Sage Foundation.

Schubert, Christian (2015). Opportunity and preference learning. *Economics and Philosophy* 31: 275–295.

Sugden, Robert (1985). Liberty, preference and choice. *Economics and Philosophy*, 1: 213–229.

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

Sugden, Robert (2006). Hume's non-instrumental and non-propositional decision theory. *Economics and Philosophy* 22: 365–391.

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

Sunstein, Cass and Richard Thaler (2003). Libertarian paternalism is not an oxymoron. *University of Chicago Law Review* 70: 1159–1202.

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

Wicksteed, Philip (1910). *The Common Sense of Political Economy*. London: Macmillan.