Economics & Philosophy (2021), **37**, 304–327
doi: 10.1017/S0266267120000243

CAMBRIDGE UNIVERSITY PRESS

BOOK REVIEWS

Humanomics: Moral Sentiments and the Wealth of Nations for the Twenty-First Century, Vernon L. Smith and Bart J. Wilson. Cambridge University Press, 2019, xx + 215 pages.
doi:10.1017/S0266267120000243

Since winning a Nobel Memorial Prize in 2002 for his experimental research on market institutions, Vernon Smith[1] has taken a new direction in his research and writing – reconstructing ideas from what he has called the 'Scottish-Hayekian' tradition of classical economics and showing their continuing relevance. In his first major work on this theme, he re-interpreted his own experimental research as a programme of empirical testing of Friedrich Hayek's account of the market as a discovery process (Smith 2008). My guess is that, for many of Hayek's disciples in the Austrian tradition of economics, the idea of Hayek as an empirical scientist proposing refutable hypotheses was too shocking to be taken seriously; but I think Vernon Smith's argument was right. This new book, co-authored by Bart Wilson, marks a new stage in Vernon Smith's reinvention of himself as a classical economist.

In a preface which sometimes sounds like a memoir of a conversion experience, Vernon Smith and Wilson (SW) tell us that it is only quite recently that they have come to understand the true meaning of Adam Smith's *Theory of Moral Sentiments* (*TMS*) (Smith 1759 [1976]) and to realize the limitations of their previous thinking, in which they were handicapped by having been educated in neoclassical economics. *Humanomics* is their attempt to explain this meaning to modern economists. (Despite the subtitle of their book, they say relatively little about *Wealth of Nations* (*WN*).) There are some born-again undertones in the reverence with which SW treat the text of *TMS*. Apart from some Google Ngrams of historical word frequencies, they do not consider *TMS* in its historical context. Nor do they countenance the possibility that there might be tensions in Smith's reasoning. They do little to relate their ideas to the various ways in which, over recent decades, other economists have interpreted and developed ideas from *TMS*. But, although these features of *Humanomics* may be off-putting to some Smith scholars, SW have important things to say about the relevance of Smith's ideas to behavioural economics.

SW's starting point is in the experimental economics of the 1980s and 1990s, which had produced 'two disparate collections of evidence'. One programme of work – the programme in which Vernon Smith had made his name – had investigated behaviour in laboratory markets and had found that this was well explained by the hypothesis that buyers and sellers act on self-interest. But another programme had found that 'anonymously paired people were predominantly caring, other-regarding, interdependent actors in the personal, social exchange context of

[1]To avoid confusion, I will attach 'Vernon' to 'Smith' whenever I refer to the co-author of *Humanomics*, and use 'Smith' to refer to Adam Smith.

trust games in the laboratory' (xiii). For reasons I will explain later, SW are not particularly moved by the findings of experiments on ultimatum games, but they are deeply impressed by Joyce Berg *et al.*'s (1995) 'unpredicted and unpredictable' discovery of trusting and trustworthy behaviour in laboratory experiments (109). (Here, SW are perhaps projecting Vernon Smith's personal shock onto the economics profession as a whole: some of us had been studying and trying to explain non-self-interested economic behaviour since the early 1980s.) It seems that Vernon Smith initially thought that he could 'uncover a crack' in Berg *et al.*'s findings, but became convinced of their external validity after running his own experiment (112). The apparent contrast between these two bodies of evidence sets up SW's research questions:

> Why do people respond trustworthily to trusting actions when it is a convenient opportunity to gain from another's largesse, as she will never know your identity? And why, when you go to the clothing store or the supermarket or Amazon, do you show so little regard for helping them by buying the highest marked-up items? (xiv)

SW undertake to answer these questions by using the theory they have found in *TMS*.

Appealing to an idea of Hayek's, SW say that all of us must 'learn to live in two sorts of world at once' – the worlds of 'sociality' and 'economy'. They have coined the neologism 'humanomics' to describe 'the very *human* problem of simultaneously living in these two worlds, the personal social and the impersonal economic'. Their thesis is that Smith's work 'enables us to understand these two worlds as one'; it 'seamlessly connects the two in a unified social and ethical science of human beings' (2). Given this claim about seamlessness, and given that *Humanomics* is primarily about *TMS*, it is reasonable to assume that SW do not mean that *TMS* and *WN* are complementary works dealing with the separate domains of sociality and economy. I take them to mean that *TMS* provides the fundamentals of a unified science and that *WN* applies or extends this science to economics.

The central chapters of *Humanomics* present a reconstruction of what SW see as the core theoretical content of *TMS*, formulated so that it yields predictions that can be tested in laboratory experiments. SW begin by sharply differentiating their Smithian approach from that of the current literature on social preferences. They emphasize two distinctions. The first is that social preference theories are ultimately theories of *behaviour*, viewed in a third-person perspective. SW accept the validity of a third-person approach for 'organizing consumer behavior in markets', but reject it as the basis for a theory of 'personal social interactions': in the latter case, they say, we need a theory of human *feelings*. In a striking analogy with Newtonian physics, SW describe *TMS* as a 'groundbreaking theory of gravity' that uses the 'hidden forces of feeling' to unify economy and sociality (41–2); 'moral sentiments are the gravity of the social universe' (46). Whether or not Smith was the Newton of the science of human feelings, SW are right to distinguish between theories of sentiment and theories of behaviour.

I was less convinced by SW's second distinction. The bedrock axiom of SW's reconstruction – they call it 'Axiom 0' – is 'common knowledge of non-satiation',

https://doi.org/10.1017/S0266267120000243 Published online by Cambridge University Press

with non-satiation defined as 'for each person more of a valued item (money or fungible goods) is better, and less of it is worse', and glossed as 'man is naturally self-loving' (69). SW argue that without a primitive concept of self-love, it would not be possible to define the moral categories of benefiting and harming others (10). So far, so good, but SW want to claim this as a distinguishing feature of Smith's theory. They criticize social preference theory for assuming that we human beings are 'other-regarding because we reductively prefer to be social' (11) – an assumption that they call 'logically circular' (xiv). But in social preference modelling it is standard practice to distinguish between 'material payoff' (effectively, a measure of quantities of some universally valued good) and 'utility' (the function that individuals are assumed to maximize). Thus, for example, a person's 'kindness' can be represented as a preference for increasing another person's material payoff. This way of theorizing is no more circular than the use of historical data about people's purchases to estimate demand functions.

The key concepts in SW's reconstruction are *beneficence*, *gratitude* and *reward* and their negative correlates *injustice*, *resentment* and *punishment*. A person's action is beneficent if it intentionally confers a benefit on another person; it is unjust if it intentionally confers a harm. Beneficent actions induce the sentiment of gratitude in the person who is benefited, and are judged both by that person and by neutral observers to deserve reward. Unjust actions induce resentment and are judged to deserve punishment.

It is essential for this theory that there is a reference point to mark the boundary between benefit and harm. SW describe this reference point as 'a normal baseline condition' (73) and suggest that it depends on the 'circumstances or context of an action' (78), but do not specify these ideas in any detail. The reference point is crucial for two reasons. First, the theory postulates an asymmetry between gains and losses: SW credit Smith with anticipating the modern concept of loss aversion (75–6, 200). Thus, the mechanism of beneficence–gratitude–reward has less psychological force than that of injustice–resentment–punishment. Second, according to the theory, an absence of beneficence does not induce resentment and is not judged to deserve punishment, and an absence of injustice does not induce gratitude and is not judged to deserve reward. Given the importance of the reference point, the imprecision of SW's concept of a 'normal baseline' weakens their claims about the explanatory power of their Smithian theory.

Using this theory, SW interpret Berg *et al.*'s findings about behaviour in trust games as evidence of the beneficence–gratitude–reward mechanism. They discuss a game in which the first mover (Player 1) can choose whether to *end* the game with each player getting \$10, or to *invest* his \$10. If he invests, the investment produces \$40 in the hands of Player 2. Player 2 then chooses whether to *keep* the whole \$40 or to *return* \$15 to Player 1, retaining the other \$25. SW interpret *invest* as beneficent, since (relative to *end*) it clearly benefits Player 2. Player 2 feels gratitude, and judges that Player 1 deserves reward. She rewards him by choosing *return* (112–19).

This analysis of trustworthiness is not very different from analyses in the social preference literature that assume preferences for reciprocating kindness. SW diverge more sharply from that literature in their analysis of ultimatum games. SW discuss an ultimatum game in which the Proposer chooses between two distributions of

https://doi.org/10.1017/S0266267120000243 Published online by Cambridge University Press

$24 – the *equal* division ($12, $12) and an *unequal* division ($2, $22) that favours her. The Responder then chooses whether to *accept* the proposal or to *reject* it, in which case both receive nothing. In experiments, highly unequal divisions are often rejected. The usual explanation is that Responders have non-self-interested preferences for 'fair' distributions. SW's explanation is that Responders are engaging in 'involuntary extortion' (135–8): they are using the threat of rejection to induce larger transfers from Proposers. On this interpretation, the evidence from ultimatum games disconfirms the predictions of theories of subgame perfection (Responders carry out 'non-credible' threats), but it does not disconfirm the assumption that players act on self-interest.

SW report an interesting experiment of their own, using the ultimatum game I have described but with an additional opening move which allows the Responder to choose whether to play the game (*in*) or not (*out*); if she chooses *out*, each player gets $1. In this game, almost all Responders choose *in*, many Proposers choose *unequal*, and a large majority of those offers are accepted. In explanation, SW say only that the players are 'voluntarily engaging in an interaction with gains from exchange' (138). The suggestion seems to be that, because the Responder entered the game voluntarily and because both players are better off with the *unequal* outcome than if the game had not been played, the Proposer's choice of *unequal* does not confer a harm on the Responder and so does not deserve punishment. But firming up this suggestion would require a definition of the 'normal baseline'.

If SW are to deliver on their promise to explain why shoppers do not go out of their way to increase the incomes of supermarket owners, they need to explain what makes retail markets different from experimental trust games. SW are ambivalent about this. They endorse the famous sentence in *WN* in which Smith says that it is not from the benevolence of the butcher, the brewer or the baker that we expect our dinner, but from their regard to their own interest; they merely note that for Smith, self-interest in the market is morally constrained by the laws of justice (4). But by putting (say) bottles of beer on his shelves, the supermarket owner is providing the shopper with an opportunity to benefit both of them by buying something to drink with her dinner. By choosing *invest* in the trust game (which in its anonymous laboratory setting is at least as impersonal as the supermarket), Player 1 is providing Player 2 with an opportunity to benefit both of them by choosing *return*. If Player 1 is beneficent, isn't the same true of the supermarket owner? And if the supermarket owner is beneficent, why don't shoppers feel gratitude? Or if they do, why does paying the posted price express exactly the right degree of gratitude?

SW sometimes say that market transactions are beneficent on both sides. Discussing Smith's treatment of market transactions in *WN*, they say that these are 'extensions of human sociality' in the sense of *TMS*: 'Each [party to a market transaction] feels net gratitude in receiving more in value than they give up. *All such trades are an exchange of gifts in the beneficence sense, that each has to give in order to receive*' (206, italics in original). But in discussing retail markets, they note that 'almost all retail consumer goods are sold at fixed, posted, take-it-or-leave it prices', and then say 'there is simply no intentional beneficence to assess in a seller's take-it-or-leave-it offer' (135, 139). But why not? If I send a donation to a charity in the form of a cheque, does the fact the charity can choose whether or not to cash it mean that I am not intentionally beneficent?

It seems that SW are having difficulty in attributing a moral status to market transactions that is consistent with their reconstruction of *TMS*. Part of the problem, I think, can be traced to their treatment of 'injustice' as the negative correlate of 'beneficence'. In Smith's work, as in 18th-century writing more generally, 'justice' has a very restricted meaning. In *TMS*, Smith sets out what seems to be intended as an exhaustive classification of the laws of justice:

> The most sacred laws of justice . . . are the laws which guard the life and person of our neighbour; the next are those which guard his property and possessions; and last of all come those which guard what are called his personal rights, or what is due to him from the promises of others. (*TMS* Part II, Section ii, Chapter 2)

Even the obligation of parents to maintain their children – surely a normal baseline condition in SW's sense – is a 'duty of beneficence' and not a demand of justice, even though it can properly be enforced by civil magistrates (*TMS* Part II, Section ii, Chapter 1). For Smith, 'natural justice' is a (more or less) well-defined and commonly understood code of context-independent moral rules, grounded on a conception of property rights. Many of SW's difficulties in specifying baseline conditions for laboratory experiments reflect the fact that subjects' property rights within experiments are often ill-defined. (For example, whether or not the behaviour of Responders in ultimatum games is extortion hinges on which player is considered to be the owner of the money that is being distributed – money that in reality is owned by the experimenter.)

I think there is also a deeper problem, which reflects a limitation of Smith's theory rather than of SW's reconstruction of it. In *TMS*, Smith is trying to explain an existing system of moral rules. Since he wants to explain this as the result of an invisible-hand mechanism driven by the natural psychological forces of sympathy and fellow-feeling, he is *looking for* a seamless explanation of the whole system. But, on his understanding of the rules of that system, those rules make a categorical distinction between justice and beneficence. As Luigino Bruni and I have argued, there is a morality of *cooperation for mutual benefit* that does not fit easily into Smith's binary scheme. To say that two individuals – perhaps the shopkeeper and the customer, or the two players in a trust game – intend that their relationship is mutually beneficial is to say more than that each is pursuing his self-interest subject to the laws of justice, but it is not to say that each is motivated by benevolence towards the other, or by gratitude for the other's benevolence to him (Bruni and Sugden 2008; Sugden 2018).

SW are surely right to say that *TMS* is a work for all time (200). But we should not read it as if it were written on the golden plates that were supposedly shown to another of Vernon Smith's namesakes.

Robert Sugden
University of East Anglia
Email: r.sugden@uea.ac.uk

References

Berg J.E., J.W. Dickhaut and K. McCabe 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* **10**, 122–142.

Bruni L. and R. Sugden 2008. Fraternity: why the market need not be a morally free zone. *Economics and Philosophy* **24**, 35–64.

Smith A. 1759 [1976]. *The Theory of Moral Sentiments*. Oxford: Oxford University Press.

Smith V.L. 2008. *Rationality in Economics: Constructivist and Ecological Forms*. Cambridge: Cambridge University Press.

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

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, choice under uncertainty, the foundations of decision and game theory, the methodology of economics, and the evolution of social conventions.

How We Cooperate, John E. Roemer. Yale University Press, 2019, 248 pages.
doi:10.1017/S0266267120000267

As I write this, we are living in strange times, through the COVID-19 pandemic and lockdown. The newspapers are filled with headlines about R numbers, excess mortality, and saving the economy versus saving lives. But some things about the situation have a comforting academic familiarity: we are living in a giant public goods game. One person leaving the house is of no danger to anyone, but if everyone goes outside there is a crowd and the virus may spread. Or perhaps it is less comforting to think of oneself as lab rat, in a permanent experiment aiming to achieve the optimal amount of social contact. Nevertheless, people have been astoundingly compliant with social isolation – surprisingly compliant according to standard economics, although less surprising to behavioural economists and researchers in other disciplines.

The messaging to encourage compliance has been diverse. One prominent message is to 'protect yourself and your family'. Alternatively, the UK Secretary of State for Health and Social Care has been exhorting people to 'do their duty'. The first of these messages expands self-interest to include pro-social motivations, the second appeals to moral principles. Roemer's theory in *How We Cooperate* falls firmly into the second camp. He objects to the approach taken in behavioural economics, of modifying classical self-regarding preferences to incorporate other-regarding motivations, such as altruism or fairness. He argues that, instead of introducing non-standard preferences, we should introduce non-standard optimization, inspired by Kant's categorical imperative, to act only in ways that one can wish would become a universal law. Roemer introduces a new equilibrium concept and shows how it can supply general microfoundations for cooperative behaviour, in both simple games and whole economies.

Roemer motivates his approach by lamenting the deficiencies of behavioural economics and it is worth listing those here, since they also provide a yardstick