

## CLASSICAL LIBERALISM AND BEHAVIOURAL ECONOMICS: CONTEMPORARY RESTATEMENTS OF CONTRACTARIANISM AND EVOLUTIONISM.

DOI: 10.7413/18281567295

**di Giacomo Brioni**

Università degli Studi di Pisa

### *Abstract*

This paper examines how traditional justifications of market liberalism have been updated to face the behavioural shift in economics. This is achieved by comparing the works of Robert Sugden and Vernon L. Smith, interpreted respectively as restatements of contractarianism and evolutionism. It is argued that, while these authors manage to elaborate market-oriented arguments that are based on a more realistic picture of human agency, they underplay the role of conflict and group interests in political relations.

**Keywords:** liberalism, behavioural economics, contractarianism, evolutionism, market.

### **1. Introduction**

The break represented by classical liberalism in the history of political thought can be summarised as follows. Liberalism attempted to provide a foundation for political coexistence that was not based on transcendent or absolute criteria, like the will of the deity or the unquestioned power of a monarch, but only on rational agreement between free individuals. The intuition that drove this effort was that political power was at the same time a necessary tool for conflict regulation and a force to be kept in check through constitutions and general rules<sup>1</sup>. The scope of politics was to be limited to the provision

---

<sup>1</sup> As an introduction to liberalism, it is still useful to refer to F.A. von Hayek, *Liberalism*, in *New Studies in Philosophy, Politics, Economics and the History of Ideas*, Routledge and Kegan Paul, London and Henley 1978, pp. 119-51 (originally written in 1973 as an entry for the Italian Treccani Encyclopedia).

of justice and other essential public goods, under the persuasion that human flourishing is achieved only when people are left free to pursue their own ends as they see fit. Then, with the birth of political economy it became clear that the market was the kind of social order that came closest to realise this ideal in practice, since it spontaneously channelled individual efforts towards collective wealth and prosperity.

The classical liberal persuasion against political interference in people's lives is captured in the anti-paternalist principle articulated by John Stuart Mill in *On Liberty*: «the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant»<sup>2</sup>. The famous debates of the 20<sup>th</sup> Century – both inside liberal theory (e.g. left-wing liberalism vs. libertarianism) and in the confrontation between liberalism and socialism – about the opposition between liberty and equality and the virtues of the market can be seen as critical notes on Mill's principle<sup>3</sup>. Does “harm” include only intentional actions, or can it be extended to external effects such as economic dispossession, loss of opportunity, and vulnerability to risk? If so, need the role of the state be reconfigured to deal with these matters, thus widening the reach of collective action? In order to tackle these questions, people like Friedrich A. von Hayek, James M. Buchanan and others felt the urge to restate the coordinates of liberalism. They held that, while complete *laissez-faire* was not a viable option, politics could be even more inefficient in solving public problems, when not overtly dangerous for social progress.

More recently, the focus of the discussion shifted from the social effects of economic action to the problem of individual rationality. The rise of behavioural economics put into question the model of agency upheld by neoclassical theory, which often provided the backbone for market-oriented arguments. Behavioural research showed that people are not perfect reasoners but flawed beings that regularly fall prey to cognitive biases and whose preferences are dictated by the features of the social context. Therefore, their mistakes concern not only collective wellbeing but also their ability to pursue their own good. This conclusion has motivated the construction of new normative frameworks for

---

<sup>2</sup> J.S. Mill, *On Liberty and other writings*, ed. by S. Collini, Cambridge University Press, Cambridge 1989 (1859), p. 13.

<sup>3</sup> I borrow the use of Mill's harm principle as a starting point for discussing liberalism's transformations from A. de Jasay, *Choice, Contract, Consent: A Restatement of Liberalism*, Institute of Economic Affairs, London 1991, pp. 22-32.

economic policy, such as libertarian paternalism, that aim to regulate both the “inward” and “outward” dimensions of human choice.

In what follows, I will examine how traditional justifications of the market have been remodelled to account for this behavioural shift and to oppose its more interventionist bent. I will consider side by side the political theories proposed by Robert Sugden and Vernon L. Smith, treating them as systematic reformulations of contractarianism on the one hand, and of evolutionism on the other. I will argue that these theories succeed in grounding the case for market institutions on our current evidence about real-world economic behaviour. However, it will be shown that they are weak to objections from perspectives that privilege political conflict and bargaining as the driving force behind social organisation<sup>4</sup>.

## 2. Rebuilding contractarianism: Robert Sugden’s principle of mutual advantage

In liberal political theory, contractarianism is often considered in opposition to utilitarianism. The classic presentation of this *querelle* is offered in John Rawls’ *A Theory of Justice*, where utilitarianism is accused of obfuscating the differences between individual conceptions of the good. At its core, utilitarianism consists in the idea that the complexity of human wellbeing can be reduced to a single *numeraire* and that the highest aggregate of the latter should be the ultimate goal of politics. This stands in sheer contrast with liberalism’s emphasis on pluralism and on each individual’s freedom to pursue their happiness as they understand it<sup>5</sup>. Contractarianism, on the other hand, manages to avoid this difficulty by shifting the problem of justice to the choice of general (i.e. constitutional) rules. According to Rawls, the agents under the «veil of ignorance» – that is, who do not know what their

---

<sup>4</sup> The basis for this work can be found in two previous papers of mine: G. Brioni, *L’economia politica costituzionale di fronte al paternalismo libertario. Su The Community of Advantage di Robert Sugden*, in P. Barrotta, A. Masala (eds.), *Algoritmi della scelta. Etica, politica e identità nell’epoca dell’intelligenza artificiale*, Pisa University Press, Pisa 2025, pp. 197-220; Id., *Ripensare i presupposti della teoria delle decisioni attraverso Adam Smith: la prospettiva della Mainline Economics e di Vernon L. Smith*, «Rivista di Filosofia», 1/2023, pp. 91-116. In the present article, however, Sugden’s and Smith’s models are compared and discussed under a different framework, that is, the development of liberalism in the face of new intellectual challenges, such as the one coming from the behavioural strand. Moreover, both paradigms are brought under the scrutiny of the “priority-of-the-political” perspective – a critical reading that is almost completely absent in the economics debate about these topics. It should be noted that Sugden himself has recently published a paper which considers side by side contractarianism and evolutionism as alternative justifications of the market but does not compare explicitly their behavioural restatements – see R. Sugden, *Hayek, Buchanan and the Justification of the Market*, in «Kyklos», 77/2024, pp. 1069-1083.

<sup>5</sup> J. Rawls, *A Theory of Justice*, The Belknap Press, Cambridge (MA) 1971, pp. 22-27.

interests will be when actual life in society begins – will agree upon a system of rules that does not try to maximise any specific conception of the good but allows for the greatest possible liberty in that matter<sup>6</sup>.

Another version of this argument was proposed in the field of political economy by James M. Buchanan, one of the founders of public choice theory and constitutional economics. The whole of Buchanan's intellectual enterprise is built on a critique of 20<sup>th</sup> Century welfare economics, which can be described as a direct application of utilitarian philosophy. Authors like Francis Y. Edgeworth and Arthur C. Pigou had looked at the fiscal policy-maker as a disinterested entity, trying to maximise social welfare by exploiting such levers as the decreasing marginal utility of income<sup>7</sup>. According to Buchanan, this «benevolent despot» model reflects a misunderstanding of the nature of the political world – which consists of no collective entities such as the “the state”, but of self-interested individuals – and of the normative criteria that should be applied to a democratic polity<sup>8</sup>. Collective rules emerge from endogenous bargaining, and the success of the interaction is measured by how the agents themselves judge their own advantage after the fact. To Buchanan, this principle is embodied in market relations, and the only way to extend it to politics is to distinguish the stage of constitutional decision from the one of day-to-day choices. When confronted with the former, individuals are more likely to reach unanimity (i.e. Pareto-optimality) because they are uncertain about their position in future interactions. Buchanan holds that constituents will negotiate rules of the game which are not biased in favour of a specific subset of society, or which at least try to minimise rent-seeking opportunities<sup>9</sup>.

In his 2018 book *The Community of Advantage*, Robert Sugden seeks to bring Buchanan's perspective up to date with the behavioural revolution in contemporary welfare economics. Sugden's enterprise is motivated by the fact that such shift has been largely dominated by thinkers who draw a connection between the findings of behavioural studies and a soft-interventionist stance in economic policy. The

---

<sup>6</sup> *Ivi*, ch. 3.

<sup>7</sup> See R.E. Wagner, *James Buchanan and Liberal Political Economy: A Rational Reconstruction*, Lexington Books, Lanham 2017, pp. 28 ff.

<sup>8</sup> J.M. Buchanan, *The Pure Theory of Government Finance: A Suggested Approach*, in «Journal of Political Economy», 57, 6/1949, pp. 496-505.

<sup>9</sup> See e.g. J.M. Buchanan – G. Tullock, *The Calculus of Consent*, Liberty Fund, Indianapolis 1999 (1962); G. Brennan, J.M. Buchanan, *The Reason of Rules: Constitutional Political Economy*, Cambridge, Cambridge University Press 1985.

behavioural approach takes issue with the neoclassical model of the rational agent, which endows people with almost inhuman cognitive powers, accompanied by flawless execution of plans under perfect and complete information. In this model, preferences are treated as given, meaning that the environment inhabited by the agent plays no role in their shaping. Behavioural studies, often by way of lab experiments, show that the exact opposite is true: human choices are almost always context-dependent. Since people have limited abilities, time and attention, their decisions are often impulsive rather than based on rational reflection. Moreover, the environment they face is usually set up (whether by a single mind or by the concurrency of independent actions) in an asymmetrical way, that is, it makes certain options stand out more than others<sup>10</sup>. This means that most of the time the preferred course of action deviates from the “optimal” or “rational” one. Therefore, people should be steered – “nudged” – towards the latter by having their choice environments reorganised by a well-meaning policy maker. This is the thrust of Thaler and Sunstein’s libertarian paternalism: this form of intervention is not directly coercive but allows people to pursue their own good «*as judged by themselves*»<sup>11</sup>.

It would seem that libertarian paternalism has succeeded in reconciling utilitarianism with liberalism. On the one hand, it is utilitarian because it deems acceptable to exert an external influence on people’s choices. On the other, it is liberal because it doesn’t operate under an aggregate conception of social welfare: it leaves people in charge of determining their own objectives and plans. According to Sugden, however, the semblance of harmony between these two philosophical positions is shattered once one pays attention to the libertarian paternalist’s depiction of the acting self. Sugden’s objections can be summarised as a three-stage argument: (a) libertarian paternalism is based on a dualistic anthropology which distinguishes between a rational and an irrational part in decision-making; (b) this dualism contradicts the behavioural principle of context-dependency of preferences, thus revealing an ungrounded normative preference for the rational agent; (c) therefore, libertarian paternalism falls back on the benevolent despot model and fails to further individual autonomy. Let’s tackle these stages in order.

---

<sup>10</sup> See the literature discussed in R. Sugden, *The Community of Advantage: A Behavioural Economist’s Defence of the Market*, Oxford University Press, Oxford 2018, pp. 7-13. For a comprehensive view of the field see N. Wilkinson – M. Klaes, *An Introduction to Behavioral Economics (Third Edition)*, Bloomsbury Academic, London 2018.

<sup>11</sup> R.H. Thaler – C.R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Yale University Press, New Haven 2008, p. 5.

As noted above, behavioural economics seeks to represent people as they act in real-life situations – «Humans», rather than «Econs», as Thaler and Sunstein would say<sup>12</sup>. Human action is characterised by various «reasoning imperfections», such as «limitations of attention, information, cognitive abilities, or self-control»<sup>13</sup>. The latter is of special importance, because it allows libertarian paternalists to transition from positive to normative theory. They assume that the human condition is one of *akrasia*: a violation of a self-imposed code of conduct due to weakness of the will<sup>14</sup>. The role of the policy maker is to employ nudging architectures in order to recover this original act of self-constitution or to uphold «the preferences that the relevant individual would have revealed»<sup>15</sup> in the absence of reasoning errors. In other words, libertarian paternalists assume that the error-laden «psychological shell» of the person hides an «inner rational agent»<sup>16</sup>. The executive ineptitude of the former inevitably distorts the “correct” commands of the latter.

However, how can this dualist ontology obtain? The essence of behavioural economics lies precisely in its refutation of what Sugden calls «integrated preferences», that is, preferences that are «stable, context-independent and internally consistent»<sup>17</sup>. Context-dependency means that no such autonomous set of preferences can be identified: the economic agent is a situational entity, whose utility function shifts with changes in the decision-making environment. On the other hand, libertarian paternalism seems to «construct two utility functions for an individual – a behavioural utility function that describes the individual’s actual behaviour, and a hypothetical utility function in which the ‘behavioural’ parameters of the behavioural function are given default values», where the latter function is «a representation of the individual’s latent preferences»<sup>18</sup>. The problem is that these latent preferences have no «empirical content». They represent the clumsy attempt to smuggle the old model of the rational agent into the new behavioural framework<sup>19</sup>. Of course, there is no *a priori* advice

---

<sup>12</sup> *Ivi*, pp. 6-8.

<sup>13</sup> R. Sugden, *The Community of Advantage*, cit., p. 55.

<sup>14</sup> See *ivi*, pp. 79-82.

<sup>15</sup> *Ivi*, p. 55.

<sup>16</sup> *Ivi*, p. 65.

<sup>17</sup> *Ivi*, p. 7.

<sup>18</sup> *Ivi*, p. 60.

<sup>19</sup> *Ivi*, pp. 68-72.

against using the neoclassical model as an external normative criterion but, in that case, it would be necessary to forego the connection between libertarian paternalism and behavioural economics.

Sugden exploits this tension to drive his argument home. Normative economics needs a behavioural foundation if its objective is to construct institutions that are charitable to the way people act in the real world. The problem with libertarian paternalism is that it interprets the behavioural evidence in a strictly *negative* way: rather than trying to derive an endogenous standard of judgement from people's actual decision-making, the libertarian paternalist prefers to criticise it by contrasting it with a standard of conduct that is manifestly idealised. This "positioning" issue is crucial, because it reveals better than anything else the utilitarian character of libertarian paternalism. Sugden, in agreement with the rest of the liberal tradition, argues that the primary addressee of policy making should be the citizen *qua* individual source of value. As Sugden puts it: «I am entitled to treat my own judgements as authoritative, not because I believe they would be endorsed by an impartial spectator, but because I am the author of my own life». On the contrary, libertarian paternalism looks at policies from an abstract point outside society, an equivalent of "social welfare", the "impartial spectator", and so on. However, such an aseptic stance has little meaning in politics, which consists of flesh-and-blood decisions. There is always *someone* who interprets and enacts the content of a notion like "social welfare", and this is the actual policy-maker. The addressee of the libertarian paternalist, then, is not the citizen, but the policy-maker, i.e., the benevolent despot<sup>20</sup>.

After the refutation of libertarian paternalism, Sugden fleshes out his reworking of the contractarian view. Given the special care that contractarianism traditionally claims to have for individual liberty, and given the new pressure that behavioural economics puts on said claim, the economist faces a difficult challenge: she must carve out a criterion that respects autonomy without ceasing to be *normative*. Normativity implies generality and abstraction from specific instances of value, because it would be otherwise impossible (with no Platonic philosopher-king in sight) to develop collective institutions. Therefore, one needs a representation of individual interest that is both «*general with respect to individuals*» and «*general with respect to applications*»<sup>21</sup>.

---

<sup>20</sup> *Ivi*, p. 20.

<sup>21</sup> *Ivi*, pp. 83-84.

Sugden responds to this challenge with the «criterion of individual opportunity», whereby «the more opportunity a person has, the more effectively his preferences will be satisfied»<sup>22</sup>. This conception circumvents the problem of choosing a specific theory of preference generation because it applies both to the neoclassical and the behavioural view. It doesn't matter whether a person always acts on strict rational choice terms or whether she is influenced by the context: an increase in her opportunity set will always go to her advantage. However, this solution works especially well in combination with behavioural economics. The latter, according to Sugden, amounts to an image of the self – what he calls the «continuing person»<sup>23</sup> – that assigns equal substance to every choice situation, even if it's incoherent with past plans or with some abstract ideal of integrated preferences. Opportunity maximisation becomes even more valuable in this case, because it applies to the intertemporal flourishing of the person. The continuing person consists of the sum of her choices over time. «In this perspective, it becomes clear how the continuing person can value the absence of constraints on her present and future actions»<sup>24</sup>. It could be argued that maximum opportunity is the behavioural equivalent of a Rawlsian principle of justice: it is what people would agree on if they undertook a content-independent reflection about potential states of the world (but without the need for the original position construct).

In Sugden's view, the opportunity criterion cascades into a justification of the market order. The gist of Sugden's reasoning, which formalises the Smithian idea of the invisible hand, is that «people *collectively* can expect the market to provide them with a rich array of opportunities to transact with one another on terms they might find *mutually* acceptable»<sup>25</sup>. The generality requirement is satisfied by the fact that a market is about transactions, the content of which need not be specified. Single preference orderings do not enter into the picture, since «in assessing a person's opportunities, we consider what she *might* conceivably want to choose, not what, given her preferences, she *does* want to choose»<sup>26</sup>. Moreover, market transactions take place only if they leave each party better off (the Pareto criterion), a fact which underlines a strong interdependency relation between individual

---

<sup>22</sup> *Ivi*, p. 85.

<sup>23</sup> *Ivi*, pp. 100-105.

<sup>24</sup> *Ivi*, p. 105.

<sup>25</sup> *Ivi*, p. 109.

<sup>26</sup> *Ivi*, p. 110.

opportunity sets. In the absence of political or material constraints, the market maximises the possible range of transactions through the entrepreneurial activity of profit-seekers, who are unlikely to leave untapped gains-from-trade on the table if they discover a more efficient use of a resource (that is, a use that relates better to consumer demand)<sup>27</sup>.

Sugden, however, does not advocate for an unregulated market. He argues that there are contractarian reasons to intervene politically to solve externality or public goods problems, especially when solutions are conceived of on a market-based structure of incentives. The bottom line is that, whenever possible, opportunities for transactions should not be restricted by state intervention. A more insidious difficulty arises when contractarianism confronts the issue of redistribution, which impacts what Sugden, following Rawls, calls the «psychological stability» of a well-ordered society<sup>28</sup>. In short, people comply with the basic norms of a polity only when they feel they can derive an advantage from the terms of the original contract. The problem is that there is an unbridgeable gap between the logic of the market, which fosters growth at the price of inequality, and the intuitive link many people establish between stability and a fair (i.e. more equal) distribution of resources. Sugden argues that the latter ideal cannot be viable unless one is prepared to shrink the potential array of transactions. However, he also recommends general schemes of redistribution in order to contain the risk of dispossession that market activity entails.

Many scholars, chief among them Francesco Guala<sup>29</sup>, have objected to Sugden's reasoning in these matters, saying that a truly behavioural political theory should take into account the possibility of an extensive political bargaining activity side by side with the market. I would argue that the limitations of Sugden's proposal lie with the structural shortcomings of contractarianism itself<sup>30</sup>. As we have seen, mutual advantage is chosen as the guiding normative principle because it satisfies the requirement of generality. Compared with the benevolent despot model, it abstains from imposing a single interpretation of interest on the whole ordering of society. However, it replicates one of the

---

<sup>27</sup> Compare *ivi*, pp. 137-139; I.M. Kirzner, *Competition and Entrepreneurship*, The University of Chicago Press, Chicago and London 1973.

<sup>28</sup> Compare R. Sugden, *The Community of Advantage*, cit., ch. 8; J. Rawls, *A Theory of Justice*, cit., p. 175.

<sup>29</sup> F. Guala, *Behavioural politics: social insurance and mutual benefit in Robert Sugden's The Community of Advantage*, in «International Review of Economics», 68/2020, pp. 89-100.

<sup>30</sup> See also C.R. Sunstein, *Voluntary Agreements*, in «Journal of Economic Methodology», 28, 4/2021, pp. 401-408; G. Brennan, H. Kliemt, *Sugden's community of advantage*, in «Journal of Economic Methodology», 28, 4/2021, pp. 374-384.

mistakes of the latter model: it bases its recommendations on an unpolitical image of policy-making. The behavioural contractarian is right in pointing out that individuals are reciprocal advantage-seekers; however, she errs in neglecting the fact that people are also bound by group and kin interests, and that they aim to further or protect these interests by way of political bargaining. Political interest groups are among the consequences of an order where the expression of claims to social positions is free<sup>31</sup>.

Borrowing the term from Joseph A. Schumpeter<sup>32</sup>, I will call this the “objection from creative destruction”. The market is a mechanism that entails a constant reshuffling of distributive states of society. It destroys acquired privileges and it punishes decisions based on wrong evaluations of risk. Market exchanges embody Pareto-improvements, but at the same time entrepreneurial action involves the exploitations of asymmetries in current market conditions. Therefore, depending on the point of observation, a market order might correspond to a set of positive or zero-sum relations. Again, individuals who feel they have been wronged or left without compensation might transform profit-seeking in proper rent-seeking behaviour, deeming it worthy to invest their resources in coalition building rather than in productive activities<sup>33</sup>. This is not done in name of some general conception of advantage (although rent-seeking sometimes receives justifications in terms of ideological views), but with an eye to the specific interests of the acting group. In this view, mutual advantage gives way to the priority of the political: the idea that people would ultimately be satisfied by an extension of their opportunity set might not be as accurate a description of human agency as Sugden suggests<sup>34</sup>. Of course, there is no denying that the market might be the most efficient ordering principle, or the most conducive to general advantage. The point of the objection from creative destruction is that the

---

<sup>31</sup> Compare J.M. Buchanan – G. Tullock, *The Calculus of Consent*, cit., pp. 282-294; D.C. North – J.J. Wallis – B.R. Weingast, *Violence and Social Orders: A Framework for Interpreting Recorded Human History*, Cambridge University Press, Cambridge 2009, pp. 140-142.

<sup>32</sup> J.A. Schumpeter, *Capitalism, Socialism and Democracy*, Routledge, London and New York 2003 (1943), pp. 81-87.

<sup>33</sup> Cfr. J.M. Buchanan, *Rent Seeking and Profit Seeking*, in J.M. Buchanan – R.D. Tollison – G. Tullock (eds.), *Toward a Theory of the Rent-Seeking Society*, Texas A&M University Press, College Station 1980, pp. 3-15; G. Tullock, *The Selected Works of Gordon Tullock, Volume 5: The Rent-Seeking Society*, Liberty Fund, Indianapolis 2005.

<sup>34</sup> This objection is a variation of what Malte Dold and Mario Rizzo write in their criticism of Sugden’s position, arguing that the opportunity criterion might not reflect the agents’ self-construed image of their autonomy. M.F. Dold – M.J. Rizzo, *The limits of opportunity-only: context-dependence and agency in behavioral welfare economics*, in «Journal of Economic Methodology», 28, 4/2021, pp. 364-373.

institutional implementation of the market depends on its predominance in the political struggle between competing views of society<sup>35</sup>.

### **3. Ecological rationality and social cooperation: evolutionary liberalism in the work of Vernon L. Smith**

Along with contractarianism, evolutionism has represented the most articulate paradigm upon which to ground a justification of free market liberalism in the 20<sup>th</sup> Century. The main advocate of this approach is Friedrich A. von Hayek, who, in such works as *The Constitution of Liberty* and (most of all) *Law, Legislation, and Liberty*, developed a comprehensive theory of law and politics by generalising his earlier contributions about the dispersed nature of information in society<sup>36</sup>. According to Hayek, our rules of conduct are not the product of a single mind, but the unintended result of individual actions over time. Institutions do not prescribe or generate codes of behaviour; rather, they crystallise existing equilibria that already embody some kind of mutual advantage<sup>37</sup>. Like market prices, which reflect the tendencies in individual transactions and redistribute information to economic players, institutions encompass general relations between people (not particular cases) and work as social landmarks that signal which claims are legitimate in social interactions<sup>38</sup>.

Hayek distinguished between two kinds of order: *cosmos*, which indicates institutions born out of social evolution, and *taxis*, meaning organisations that have been consciously planned with a specific objective in mind. Complex human societies are usually a mix of *cosmos* and *taxis*, and Hayek does not discount the role of governments (the chief example of planned hierarchy) in producing public goods and services<sup>39</sup>. However, Hayek attributes the success of western civilisation to the adoption

---

<sup>35</sup> Here, I follow the lead of contemporary political realists, such as Raymond Geuss, who have emphasised that political equilibria always reflect asymmetries of power between groups and conceal the traces of previous struggles for hegemony. In this sense, there is an intimate contradiction between contractarianism's tension towards unanimous agreement in constitutional choice, and a more "embodied" political view of human interests. See R. Geuss, *Philosophy and Real Politics*, Princeton University Press, Princeton and Oxford 2008. Or, as Guala puts it, «behavioural politics boils down to this: social contracts (and the extent of redistribution) reflect bargaining power». F. Guala, *Behavioural politics*, cit., p. 99.

<sup>36</sup> See F. A. von Hayek, *Economics and Knowledge* and *The Use of Knowledge in Society*, both in *Individualism and Economic Order*, The University of Chicago Press, Chicago and London 1948, pp. 42-65 and 86-100.

<sup>37</sup> Id., *Law, Legislation, and Liberty*, The University of Chicago Press, Chicago and London 2021 (1973-1979), pp. 28-40.

<sup>38</sup> See also Id., *Degrees of Explanation*, «The British Journal for the Philosophy of Science», 6(23), 1955, pp. 209-255.

<sup>39</sup> Id., *Law, Legislation, and Liberty*, cit., ch. 2

of impersonal rules of conduct, namely private property and contract law, which he argues are best safeguarded by common law systems. Hayek states that such rules are in harmony with the nature of historical processes, which are not teleological but have always shown the evolutionary advantage of following the human propensity for exchange rather than tribal impulses<sup>40</sup>.

While Hayek is not strictly part of the behavioural revolution in economics, his theory includes a strong behavioural component, since it is meant to describe how agents act in the real world, under constraints of information scarcity and incompleteness. According to Hayek, man is in essence a «rule-following animal»<sup>41</sup>. Human patterns of conduct are not consciously chosen on every single occasion on account of a utility calculus but are to be seen as “social filters” to which individuals adhere because everyone else does. Following David Hume, Hayek holds that mutual advantage and reciprocity are meta-reasons for the existence of such filters; however, these reasons need not be made evident to actors every time<sup>42</sup>. This intuition was expanded upon by Sugden in his 1985 book *The Economics of Rights, Co-operation and Welfare*, where he uses game theory to describe rules as «evolutionarily stable strategies» that invade the population of players and become self-perpetuating conventions<sup>43</sup>. Sugden thought of this idea as a tool to disengage contractarianism (especially in the Rawls-Buchanan version) from such problematic notions as the veil of ignorance and reflective equilibrium, which, even with relaxed assumptions, involve unrealistic moral and epistemic requirements. The evolutionary perspective, on the other hand, balances rule compliance with a more realistic conception of how individual interests play out<sup>44</sup>.

However, the most meaningful reformulation of Hayek’s philosophy in a behavioural framework is found in the works of Vernon L. Smith. Smith pioneered lab experiments in economics, producing evidence about the emergence of cooperation where economic theory would have predicted zero-sum interactions, and about the validity of traditional hypotheses of market equilibrium. In his *Rationality in Economics*, Smith translates Hayek’s *taxis-cosmos* dichotomy in terms of different modes of

---

<sup>40</sup> Id., *The Fatal Conceit: The Errors of Socialism*, Routledge, London 1988, pp. 31-47.

<sup>41</sup> Id., *Law, Legislation, and Liberty*, cit., p. 28.

<sup>42</sup> See Id., *The Legal and Political Philosophy of David Hume*, in *Studies in Philosophy, Politics, and Economics*, The University of Chicago Press, Chicago-London 1967, pp. 106-121.

<sup>43</sup> R. Sugden, *The Economics of Rights, Co-operation and Welfare*, Palgrave Macmillan 2005 (1986), pp. 29-35.

<sup>44</sup> *Ivi*, pp. viii, 179, 224-225.

rationality. «Constructivist rationality» describes man's conscious effort to organise social environments around an explicit goal, whereas «ecological rationality» refers to how these efforts are filtered by experience, history, and trial-and-error processes, that is, social pathways upon which no discretionary control could be exercised<sup>45</sup>.

Now, in his work Hayek equates *taxis* to small-group interactions and *cosmos* to what he calls the «extended order» of the market, made of mostly impersonal relations. Hayek's last book, *The Fatal Conceit: The Errors of Socialism*, leaves the reader with a dilemma. On the one hand, *cosmos* and *taxis* seem to be mutually contradictory in nature, if not mutually destructive, since they involve different conceptions of justice, affection, utility, and so on: applying only self-interest to, say, family relations would bring about their dissipation, whereas running a firm on criteria of liberality as opposed to profit-seeking would soon result in bankruptcy<sup>46</sup>. On the other hand, modern man lives in both worlds at once. How can this coexistence obtain? Smith answers by going back to one of Hayek's main sources: the thought of Adam Smith<sup>47</sup>. Hayek's dilemma has the same structure as the well-known *Adam Smith Problem*, that is, the seeming inconsistency between the theories of cooperation put forward in *The Theory of Moral Sentiments* and in *The Wealth of Nations* (other-regarding vs. self-regarding types of conduct). Vernon Smith's intention, which is carried out between *Rationality in Economics* and the 2019 monograph *Humanomics: Moral Sentiments and the Wealth of Nations for the Twenty-First Century* (co-written with Bart Wilson), is to show not only that face-to-face and impersonal systems are compatible but that the behavioural structure of the latter actually *derives* from that of the former<sup>48</sup>.

Smith's argument seeks to articulate Adam Smith's guiding principles in *The Theory of Moral Sentiments*, that is, fellow-feeling and sympathy, with reference to experimental evidence from distributional interactions such as trust and ultimatum games. In the trust game, player A is given the

---

<sup>45</sup> V.L. Smith, *Rationality in Economics: Constructivist and Ecological Forms*, Cambridge University Press, Cambridge 2008, pp. 26-32, 36-41.

<sup>46</sup> F.A. von Hayek, *The Fatal Conceit*, cit., pp. 18-19.

<sup>47</sup> On Vernon Smith's reading of Adam Smith see also M.P. Paganelli, *The same face of the two Smiths: Adam Smith and Vernon Smith*, «Journal of Economic Behavior and Organization», 78, 3/2011, pp. 246-255.

<sup>48</sup> Compare V.L. Smith – B.J. Wilson, *Humanomics: Moral Sentiments and the Wealth of Nations for the Twenty-First Century*, Cambridge University Press, Cambridge 2019, pp. 3-4; V.L. Smith, *Adam Smith and Experimental Economics: Sentiments to Wealth*, in R.P. Hanley (ed.), *Adam Smith: His Life, Thought, and Legacy*, Princeton University Press, Princeton and Oxford 2016, pp. 262-280 (p. 268).

opportunity to either distribute a sum of money equally between them and player B, or to pass the responsibility to B, who can in turn decide to share with A the now multiplied amount, or to keep the whole of it. In the ultimatum game, player A must make an offer to player B about how to distribute the given amount; B can accept or refuse, but in the latter case no one gets anything. In both instances, the neoclassical model of the self-interested agent would predict no cooperation: the trust game ends immediately with A settling for the first option in fear that B will act selfishly, and the ultimatum game results in A offering to B the smallest share possible, which B will be forced to accept<sup>49</sup>.

Smith, along with a well-established literature about cooperation in repeated games<sup>50</sup>, finds out that even with stricter versions of these experiments there is a high rate of players adopting the collaborative strategy. One could be inclined to say, like many scholars have done, that self-interest must be replaced by other-regarding motives, whereby agents incorporate the other player's utility in their own utility function. Smith rejects this explanation as too outcome-oriented, and favours a reading based on a combination of self-interest and reciprocity, which he holds to be in line with the logic of *Moral Sentiments*<sup>51</sup>. According to Adam Smith, moral rules arise because people observe and judge each other's behaviour, with particular respect to feelings of gratitude and resentment. This external eye becomes progressively internalised, since each course of action is chosen on the basis of how others could react to it<sup>52</sup>. As Vernon Smith notes, «out of a desire to satisfy our social impulse, human beings self-examine their own sentiments, impartially judge their own conduct, and exercise self-command in resonance with the judgment of others»<sup>53</sup>. From this constant process of inter- and infra-subjective feedback<sup>54</sup>, a shared criterion of action emerges, as if brought about by an impersonal entity that oversees the whole of society. This is what Adam Smith calls the «impartial spectator», a

---

<sup>49</sup> V.L. Smith – B.J. Wilson, *Humanomics*, cit., ch. 8-9.

<sup>50</sup> The starting point is R. Axelrod, *The Evolution of Cooperation*, Basic Books, New York 1984.

<sup>51</sup> V.L. Smith – B.J. Wilson, *Humanomics*, cit pp. 38-41, where the relevant literature about “social preferences” is discussed.

<sup>52</sup> I won't go into detail about how Vernon Smith reads and formalises each key proposition from *The Theory of Moral Sentiments*. For the core of this analysis, see V.L. Smith – B.J. Wilson, *Humanomics*, cit., ch. 5.

<sup>53</sup> *Ivi*, p. 74. This principle articulates A. Smith, *The Theory of Moral Sentiments*, Oxford University Press, Oxford 1976 (1759-1790), III.i, pp. 109-113.

<sup>54</sup> See M. Bee, *The Pleasure of Exchange: Adam Smith's Third Kind of Self-Love*, «Journal of the History of Economic Thought», 43, 1/2021, pp. 118-140: Bee ties Smith's concept of self-love to the need for mutual recognition and reads it as the force that drives economic exchange.

concept that is often (e.g. by Rawls) associated with utilitarianism but which Vernon Smith shows to be divorced from any consideration of reciprocal or general utility. The impartial spectator is rather the result of fellow-feeling filtered through cultural evolution and the workings of social psychology<sup>55</sup>.

Gratitude and resentment are what motivates reward and punishment respectively, that is, the backbone of our intuitions about justice<sup>56</sup>. Smith highlights their relevance by devising variations of the trust and ultimatum games that incorporate options to retaliate against the other player's decisions or to reward beneficence. In this way, it is shown, for instance, that players tend to punish intentional harm but not the absence of beneficence, a conclusion that Adam Smith would have subscribed to. Moreover, experimental results back the Smithian intuition that the intensity of resentment and gratitude is differentiated and tuned to fit the particular circumstances to which these sentiments apply<sup>57</sup>. This is especially important in the case where one has to judge the right punishment. «The rules of fair play require the punishment to be appropriate – to be reasonable in fitting the resentment to the infraction. Therefore, if many first movers impartially judge the second movers' misstep, and see the punishment as excessive, they will be loath to impose such a high punishment for defection. The impartial spectator is not vindictive»<sup>58</sup>. In short, justice goes hand in hand with propriety, defined as equanimity in judgement. Moreover, the exercise of propriety in single instances of praising and blaming contributes to the development of a more general sensibility towards praiseworthiness and blameworthiness<sup>59</sup>.

The upshot of this argument is that such emergent rules of justice will apply also to *n*-people anonymous contexts, as is the case with market transactions<sup>60</sup>. In examining the passage between *Moral Sentiments* and *The Wealth of Nations*, Vernon Smith works out the connection between propriety and private property: «property in the civil order emerges gradually from our long

---

<sup>55</sup> Compare J. Rawls, *A Theory of Justice*, cit., pp. 27, 29, 33, 88-89, 184; V.L. Smith – B.J. Wilson, *Humanomics*, cit., pp. 101-103.

<sup>56</sup> E.g., V.L. Smith – B.J. Wilson, *Humanomics*, cit., pp. 71-73, 105-106.

<sup>57</sup> Compare *ivi*, pp. 149-160; V.L. Smith, *Rationality in Economics*, cit., pp. 267-272.

<sup>58</sup> V.L. Smith – B.J. Wilson, *Humanomics*, cit., p. 153, 87, which quotes A. Smith, *Moral Sentiments*, cit., II.ii.1, p. 112.

<sup>59</sup> V.L. Smith – B.J. Wilson, *Humanomics*, cit., p. 162, 77, with reference to A. Smith, *Moral Sentiments*, cit., III.i.2, p. 166.

<sup>60</sup> V.L. Smith, *Rationality in Economics*, cit., ch. 6

experience of the rules of propriety in the social order»<sup>61</sup>. Private property and contract are institutions that embody what we expect from other people's conduct: fair play consists in not violating the rights and obligations which are generated by possession and promises. Naturally, the transition from small groups to the extended order requires third-party, formal mechanisms for conflict resolution<sup>62</sup>. Such institutions, however, would not stand the test of time if they worked only on the basis of coercion. Their legitimacy lies in the evolutionary memory of cooperative interactions<sup>63</sup>.

Moreover, the model developed in *Moral Sentiments* is at work in how Adam Smith interprets the growth of specialisation and trade in the market. «Property – the universal human custom of mutually recognizing what is mine and thine – is necessary but not sufficient». What is needed is «the human “propensity to truck, barter and exchange.” Just as disinterested commerce underlies the social order, Smith's axiom is simply an extension of human sociability to interested commerce; it is “commerce” all the way up, from neighborly social exchange to the extended order of impersonal markets»<sup>64</sup>. Fellow-feeling is the principle from which both close-knit affection and market exchange spring, although at different levels of generalisation.

Like Sugden's, Vernon Smith's theory offers a plausible account of how the essential features of market society can emerge and gain stability from the independent decisions of self-interested individuals. Although Smith does not explicitly formulate his views in opposition to behavioural paternalism, the stress he puts on spontaneous reciprocity as the driving force behind cultural evolution establishes by itself an anti-paternalist interpretation of man's social conduct. Like Hayek before him, Smith warns intellectuals about the dangers of the constructivist approach, which, if applied without regard for incentives and individual expectations, threatens to destroy the very determinants of growth.

At this point, it can be useful to apply to Smith the same critical remarks levelled against Sugden. For one, does the array of experimental results examined by Smith give us a comprehensive enough

---

<sup>61</sup> Id., *Adam Smith and Experimental Economics*, cit., p. 272, which quotes A. Smith, *Moral Sentiments*, cit., II.ii.2.2, pp. 82-85.

<sup>62</sup> *Ivi*, p. 270.

<sup>63</sup> Therefore, the state is the product, not the cause, of legal constraints. See Id., *Rationality in Economics*, cit., pp. 196-198.

<sup>64</sup> V.L. Smith – B.J. Wilson, *Humanomics*, cit., pp. 14-15. Compare A. Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, Oxford University Press, Oxford 1976 (1776), I.ii, p. 25.

picture of the evolution of norms? Is the passage from propriety to property the only option, or can there be pathways to other kinds of order? If the impartial spectator is the product of cultural evolution, one is tempted to say that Smith's theory should be open to a multitude of social equilibria, determined by history, culture, individual attitudes, etc<sup>65</sup>. This problem is analogous to the one generated by the use of Schelling points as a justification for private property in problems of externality and resource use. The concept of a Schelling point indicates a convergence on a given option based on a shared understanding of its more «conspicuous» qualities. However, if two options stand out equally, it is not easy to say which will be the preferred solution. In the case of distributional issues, the conflict could take the form of equality vs. private property (both have the property of being an «extreme» solution if put on a continuum)<sup>66</sup>. The choice will be the outcome of a mix of situational elements and features of the individual players, but there is no way to predict the result *a priori*. Vernon Smith's theory is certainly more evidence-driven than intuitive speculation about Schelling points, but the problem crops up again at the meta-level of cultural boundaries.

Second, Smith seems again to eschew properly political interactions from his picture. It is true that he manages to describe how Hayekian general rules springs from personal interactions, or how even close-knit relations are related to ecological rationality. However, one is left with the impression that this is just one side of the story. I would argue that Hayek and Smith have slightly different conceptions of small-group behaviour, with Hayek stressing the logic of rent-seeking and protection of special interests, that is, how the «tribe» relates to the outside world. On the other hand, Smith takes a more inward perspective, looking at affective behaviour between self-interested individuals. In this way, however, the original Hayekian dilemma remains intact, because Smith's model does not take into account the emergence of political conflict side by side with cooperation. While this does not diminish the normative significance of ecological rationality, as the indictment against

---

<sup>65</sup> Note, however, that there is no unanimity in the Adam Smith scholarship about the interpretation of the impartial spectator. In some cases, the latter is taken to be a culture-bound concept; others argue that Smith distinguishes between imperfect and ideal forms of the impartial spectator, where the latter is used as a measure of social progress. Compare D. Golemboski, *The impartiality of Smith's spectator: The problem of parochialism and the possibility of social critique*, in «European Journal of Political Theory», 17, 2/2015, pp. 1-20; J.T. Young, *Natural Morality and the Ideal Impartial Spectator in Adam Smith*, in «International Journal of Social Economics», 19, 10-12/1992, pp. 71-82.

<sup>66</sup> Compare T. Schelling, *The Strategy of Conflict*, Harvard University Press, Cambridge (MA) 1960, ch. 3; R. Sugden, *The Economics of Rights*, cit., pp. 49-54, 95-106; B. Barry, *Political Argument*, Routledge & Kegan Paul, London 1965, pp. 319-322.

constructivism based upon Hayek's theory of knowledge still holds, it hinders the completeness of Smith's evolutionism as a properly *political* theory.

#### 4. Concluding remarks

In this paper, I have examined how traditional paradigms of market liberalism can be adapted to face the challenge of behavioural economics. Both contractarianism and evolutionism start from a concern with finding the baseline of agreement between individuals with different conceptions of the good. These theories have always sought to elaborate an endogenous theory of the state, without reference to external normative criteria. At the same time, they express a balance between description and justification, pointing to the market as a social model that embodies opportunities for mutual advantage (Sugden) and norms of reciprocity or fair play (Smith). This kind of order fits a world in which individuals, contrary to what neoclassical economics had stated, do not have stable and integrated preferences. Moreover, it reflects the fact that self-interested agents often come to cooperate with each other in situations where conflict should be expected to dominate.

While these theories fare well against corresponding re-evaluations of utilitarian or statist point of view in behavioural theory (such as libertarian paternalism), they are somewhat vulnerable to criticism from a more politics-centred perspective. In both cases, reasons for agreement on liberal principles are supposed to trump considerations in favour of rent-seeking, but real-world democracies offer ample evidence that the latter has a strong hold over ordinary political decision-making. Contractarianism and evolutionism seem to assume away, rather than tackle, the problem of conflict. A comprehensive behavioural perspective, on the other hand, should account for it in the same way it explains cooperation and market exchange<sup>67</sup>.

---

<sup>67</sup> As Guala suggests, behavioural politics, as opposed to behavioural economics, still lacks a «unifying framework». F. Guala, *Behavioural politics*, cit., p. 91.



Sesto San Giovanni (MI)  
via Monfalcone, 17/19



& Ass. AlboVersorio Edizioni  
Senago (MI)  
via Martiri di Belfiore, 11

© Metabasis.it, rivista semestrale di filosofia e comunicazione.  
Autorizzazione del Tribunale di Varese n. 893 del 23/02/2006.  
ISSN 1828-1567



Quest'opera è stata rilasciata sotto la licenza Creative Commons Attribuzione-NonCommerciale-NoOpereDerivate 2.5 Italy. Per leggere una copia della licenza visita il sito web <http://creativecommons.org/licenses/by-nc-nd/2.5/it/> o spedisci una lettera a Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.