The Contribution of Behavioural Insights to Political Economy

Malte Dold
Pomona College

and

Paul Lewis
King’s College London

Forthcoming in the Review of Behavioral Economics

Abstract
This paper is an introduction to the special issue of the Review of Behavioural Economics in honour of Shaun Hargreaves Heap. The essays contained in the special issue explore, via an engagement with Hargreaves Heap’s work, how behavioural insights can contribute to the discipline of political economy. This paper introduces the essays by summarising how they suggest that arguments and findings derived from behavioural economics can be used to address important issues in contemporary political economy.
Shaun Hargreaves Heap (hereafter, SHH), whose work in experimental and behavioural economics has inspired this special issue, read Philosophy, Politics and Economics at Oxford University, after which he completed a PhD in Economics at the University of California, Berkeley. Early research in macroeconomics (Hargreaves Heap, 1980, 1992) was accompanied by studies of issues arising from the role of rationality assumptions in economics (Hargreaves Heap 1989; Hargreaves Heap and Varoufakis 1995). Themes addressed in that work, most notably concerning the shortcomings for both explanatory and normative purposes of the instrumental account of rational action, also featured prominently in SHH’s sustained engagement, now lasting for two decades, with experimental and behavioural economics. It is methodological and normative questions arising from SHH’s work in experimental and behavioural economics that are the focus of this special issue, contributions to which explore—through reflections on SHH’s work—the relevance of insights derived from experimental and behavioural economics for contemporary political economy. Given that an interest in political economy, understood as the exploration of the interactions between economic and political processes, has long been a feature of SHH’s work (Hargreaves Heap and Hollis 1984, Hargreaves Heap 2020), this is a fitting topic for a symposium in his honour, as well as one that coheres with, and contributes to, the recent resurgence of interest in political economy in the social sciences. The special issue consists of six papers that explore the contribution of behavioural insights to political economy from multiple angles.

Peter John begins his paper on “Shaun Hargreaves Heap’s Behavioural Political Economy: Implications for Politics and Public Policy” by highlighting SHH’s early work on the shortcomings of the standard economic account of instrumentally rational conduct, as set out in his 1989 book, *Rationality in Economics* and also in his book, co-authored with Yanis Varoufakis, *Game Theory: A Critical Introduction* (Hargreaves Heap 1989; Hargreaves Heap and Varoufakis 1995). In the former, as John explains, SHH argues that “mainstream economics is ill-served by its exclusive, formal reliance on the instrumental sense of rationality”, according to which rational conduct involves the efficient use of the means at people’s disposal to satisfy a given set of well-defined preferences (Hargreaves Heap 1989: 2). Whilst acknowledging that instrumental rationality is a powerful explanatory tool, SHH argues that it cannot on its own provide an adequate understanding of various kinds of important economic phenomena, satisfactory explanations of which require that it be augmented by an alternative account of expressively rational conduct. According to the latter, a rational individual is “self-consciously reflective about his/her preferences” and seeks preferences that are “worthy of one’s self”, so that action “is part of a groping towards who to be, what ends to have” (Hargreaves Heap 1989: 5). In this way, as John observes, SHH seeks to explore “the limits of a particular form of individualism in social science: one based exclusively on the model of persons as preference satisfiers” (Hargreaves Heap and Varoufakis 1995: 2; quoted by John).
John observes that SHH subsequently built on this work by engaging in a series of wide-ranging projects in experimental economics, encompassing work on, to name but a few topics, the emergence, and subsequent consequences, of conventions about fairness in cases of strategic interaction (Hargreaves Heap and Varoufakis 2002); the relationship between group membership and trust (Hargreaves Heap and Zizzo 2009); the impact of competition and inequality on teamwork (Hargreaves Heap et al. 2015, 2021, Cavatorta et al., 2023); the influence of inequality on tribalism and zero-sum thinking (Hargreaves Heap 2020); and the scope for participatory decision-making to enhance productivity (Hargreaves Heap 2004). As John points out, the behavioural insights gleaned from these experiments, and from the wider literature in behavioural and experimental economics, have in turn informed a distinctive perspective on behavioural public policy (Hargreaves Heap 2013, 2017, 2023). SHH is sceptical about nudging, on the grounds that experimental evidence concerning the instability and context-dependence of people’s preferences indicates that there is no ‘true’ or ‘correct’ set of preferences whose satisfaction policy should target. Policy should focus instead on enhancing people’s autonomy, thereby taking what John describes as a “balanced approach” according to which,

The options are not simply to be quiet on policy in the face of anomalies (Bernheim) or be in some degree paternalist (Sunstein and Thaler). Policy could avoid taking a view on people’s true interests while still being active if it was directed at the conditions under which people acquire the sense of interest on which they act. (Hargreaves Heap 2013: 995; quoted by John).\(^1\)

As John goes on to note, this approach “puts [Hargreaves Heap] in the vanguard of current debates about behavioural public policy toward providing for more autonomous choices (see the review by Banerjee et al., 2023), as in Sunstein’s deliberative nudges (Sunstein, 2016) or deliberative thinks (John et al., 2019), or arguments for encouraging more individual capacity in making informed decisions as is the boost framework (Grüne-Yanoff & Hertwig, 2016; Hertwig, 2017), or adding a think alongside a nudge, as in the nudge+ suite of interventions (Banerjee & John, 2021, 2023).”

In “Towards a Liberal Behavioural Political Economy: The Constitutional Approach and the Role of Capable Agency”, Malte Dold and Paul Lewis also focus on SHH’s emphasis on autonomy as a core normative criterion. As noted above, SHH maintains that behavioural findings concerning the instability and context-dependence of people’s preferences indicate that behavioural normative economics should focus less on the satisfaction of people’s preferences and more on the extent to which people are able to form their preferences autonomously. As SHH writes, “it … matters (or ought to) for those with liberal instincts that whatever action people take, they should feel they own it in the sense that they have had the resources to reflect on what preferences to hold and how to act on them; that is, they should feel autonomous” (Hargreaves heap 2013: 995; quoted by Dold and Lewis). SHH understands ‘autonomy’ as individuals’ ability “to reflect as sovereign agents on what matters to them”, in particular about “what preferences to hold and how to act on them”, so that people become the “authors of who they are” (2022: 61, 2013: 995, 2023: 2; quoted by Dold and Lewis). Dold and Lewis argue that this shift in focus from preference satisfaction to autonomy is consistent with recent contributions to behavioural normative economics that emphasise the importance of agency, understood as people’s capacity to “own […] the process of behavioural change” (Banerjee et al. 2023: 1) and thereby be “the authors of their own lives” (Dold and Lewis 2023: 2).

---

\(^1\) In this passage, SHH refers to Bernheim (2009) and to Thaler and Sunstein (2008).
However, drawing on work carried out in positive psychology under the heading of ‘self-determination theory’ (SDT), Dold and Lewis argue that SHH concentrates so much on autonomy that he does not do complete justice to two other important dimensions of agency identified by SDT, namely competence (understood as the experience of being effective and successful in pursuing goals) and relatedness (conceptualised as a sense of contributing to, and feeling connected with, others) (Ryan and Deci 2000, 2006; Vansteenkiste et al. 2020). Dold and Lewis contend that people’s feelings of autonomy, competence and relatedness—in short, their feelings of agency—are underpinned by the possession of certain agentic capabilities (where the term ‘capabilities’ is used in the sense pioneered by Amartya Sen to denote the kinds of activities that people can take and the types of lives they can lead) (Sen 1985, 1999; also see Dold and Lewis 2023: 5–7). In developing SHH’s account in this way, Dold and Lewis suggest that they are advancing an account of capable agency that is not only compatible with SHH’s existing work but is also, to some extent at least, already present in it (albeit without being set out as explicitly and systematically as it might have been). Having set out this view of capable agency, Dold and Lewis then also proceed to explore the social-structural conditions, including policies, that are conducive to fostering people’s agentic capabilities and sense of agency.

In a contribution entitled “Self-realisation and Usefulness: A Critical Examination of Self-determination Theory”, Robert Sugden responds to recent attempts by economists to argue for individual autonomy as a normative benchmark in public policy discussions. Sugden is critical of the move to treat autonomy as a dimension of value in public policy, similar to other values like welfare, efficiency or distributional equality. Yet, he has an interest in this approach as “it has been presented as an alternative to my own attempts to develop a form of normative analysis that is rooted in the liberal tradition of economics but compatible with the findings of behavioural economics.” In previous work, Sugden has defended the idea of taking opportunity, in particular individuals’ opportunities to engage in voluntary transactions, as the benchmark for making normative judgements and has argued that competitive markets are the best means of providing such opportunities (Sugden 2004, 2018; McQuillin and Sugden, 2012). In this contribution, Sugden critically discusses SDT as it is often invoked by those who defend autonomy as a value (see, e.g., Hargreaves Heap 2013, 2022, Dold and Lewis 2023 and the article by Dold and Lewis in this special issue). In particular, Sugden argues that SDT’s categorisation of motivation along an intrinsic-extrinsic spectrum fails adequately to acknowledge how real individuals derive satisfaction from being useful to one another; he also believes that seeing market transaction through an SDT lens obscures their moral content as it falsely emphasizes the importance of self-realization.

In SDT, intrinsic motivation is defined as a form of motivation that is directed at an activity that a person freely engages in out of interest, for the fun or challenge involved, without the necessity of separable consequences, such as external prods, pressures, or rewards (Deci and Ryan 2000: 233). Sugden points out that Deci and Ryan treat this construct as the prototype of autonomous action. To challenge SDT, Sugden uses an example of him and his wife doing the dishes every evening. He says that he is not intrinsically motivated to do the dishes; he has no desire to dirty them and then re-wash them for the fun or challenge entailed; he also does not feel prodded or pressured into dishwashing. Sugden argues that this example reveals a limitation of SDT’s intrinsic/extrinsic categorisation of motivations since it fails to recognise that activities can have intrinsic purposes. For many activities in everyday life, such as dishwashing or mundane market activities, the purpose is to be useful to one another. And when participating in these activities, we often take on its purpose as our motivation. Sugden argues that this purpose-conditional motivation is crucially different from SDT’s concept of integrated extrinsic motivation. SDT presupposes that individuals have a natural tendency to integrate socially sanctioned norms and requests into personally endorsed values. This integration is assumed to be a matter of self-realisation and of harmonising external prods and pressures.
within a unified sense of self. Sugden states that the assumption of “a psychological desire to unify the disparate purposes of human life into a single sense of self strikes [him] as implausible – not to say narcissistic.” According to Sugden, human behaviour is inherently context-dependent; it involves taking on a purpose that is given by the context in which we find ourselves in. This context-dependence, Sugden argues, is compatible with the opportunity perspective on normative economics he has advanced but challenges both standard behavioural welfare economics and the agentic approach supported by, inter alia, Hargraves Heap (2013, 2017, 2022), Dold and Stanton (2021), and Dold and Lewis (2023). It challenges standard behavioural welfare economics because it pushes back against the psychological realism of a core of latent or true preferences against which one could classify actions as more or less erroneous. And, according to Sugden, it challenges the agentic approach with its normative emphasis on the reflective self that explicitly or implicitly assumes an ideal of an integrated sense of self against which actions can be judged as more or less autonomous. Sugden argues that the assumption of an integrated sense of self “treats the ordinary workings of human psychology as inauthentic.” In particular, he finds fault with the agentic approach’s emphasis on critical self-reflection and self-creation and the normative ideal of a society in which individuals can engage in lifelong projects of preference learning. Sugden argues that “self-creation is further from many people’s minds than the agentic approach supposes.” It should therefore not be a proper concern of public policy.

In “Inequality as a Behavioral Driver: An Inspiring Contribution to Behavioral Political Economy”, Enrique Fatas and Lina Restrepo-Plaza discuss one of SHH’s most cited contributions to behavioural political economy, namely his 2002 Economic Journal paper, co-authored with Yanis Varoufakis, on “Some Experimental Evidence on the Evolution of Discrimination, Co-Operation and Perceptions of Fairness” (Hargreaves Heap and Varoufakis 2002, hereafter HHV). Drawing on the Hawk-Dove game, HHV’s experiment revealed the emergence of a discriminatory convention based on arbitrary labels where half of the participants were randomly given red labels and the remaining half received blue labels. Subsequently, when the option to cooperate was introduced, those who were disadvantaged by the colour-based discriminatory convention tended to cooperate with each other, whereas the others did not. Advantaged subjects tended to get caught in a vicious convention and play ‘Hawk,’ which came at a significant cost to them. Disadvantaged subjects, on the other hand, were more likely to cooperate, personally profiting from this less conflictual strategy. In their contribution to this special issue, Fatas and Restrepo-Plaza’s primary goal is not to replicate HHV but rather to demonstrate how it has inspired fresh ideas for research, contributing to the emergence of a new generation of papers that concentrate on the link between inequality and behaviour – a topic of central importance for discussions in behavioural political economy and institutional design.

Inspired by HHV, Fatas and Restrepo-Plaza discuss two recent studies, one in the lab and one in the field, that explore the behavioural effects of exposure to advantageous and disadvantageous circumstances. The first study, by Eckel et al. (2022, EFK hereafter), utilizes a methodology akin to that employed in HHV (a lab experiment featuring arbitrary minimal identities). EFK examines how various forms of inequality influence the conduct of disadvantaged individuals. It demonstrates that in conditions of unequal distribution of rights and opportunities, such as in the political inequality scenario, the intergroup dynamic takes an unfortunate trajectory. Disadvantaged individuals partake in costly intergroup punishment at a large personal cost. As conflict intensifies, the escalation of intergroup punishment spirals out of control, with advantaged players unable to deescalate it through resource redistribution, resulting in significant social welfare losses. The second study discussed is a lab-in-the-field study by Restrepo-Plaza and Fatas (2022, RPF hereafter) in which the cooperative behaviour of individuals with naturally occurring identities in post-conflict Columbia (victims of conflict,
non-victims, and ex-combatants) is analysed in a social dilemma situation. Both victims of conflict and non-victims belong to similar vulnerable low-income groups, but non-victims experience slightly less disadvantage due to their higher socioeconomic status and lack of exposure to violent conflict. Through a series of one-shot games pairing participants with members of each group, RPF analyses intergroup cooperation and discrimination. The authors argue that the prosocial behaviour identified in RPF is consistent with the cooperative behaviour identified in HHV: disadvantaged individuals (victims) exhibit less discrimination towards ex-combatants than non-victims, and they treat members of other disadvantaged groups (non-victims) even more favourably than they treat themselves, as evidenced by increased cooperation when matched with a non-victim compared to when matched with a victim.

This research aligns with HHV by revealing how different (sometimes arbitrarily created) identities and economic or political opportunities lead to different behavioural outcomes. Fatas and Restrepo-Plaza contend that a crucial takeaway from this set of papers is the illustration of how disadvantaged individuals occasionally diverge from arbitrary norms, benefiting from out-of-equilibrium cooperative strategies (as seen in HHV). They tend to exhibit greater pro-sociality and diminished discriminatory biases compared to less disadvantaged participants (as demonstrated in RPF). In addition, EFK underscores the steep costs of intergroup conflict when linked to an unequal distribution of opportunities. Once initiated, even if the privileged understand that it would be in their own interest to ensure a less unequal regime, they are unable to rein it in.

In a contribution entitled “Heap-ing on Lippmann: Liberalising Behavioural Public Policy”, Adam Oliver seeks to join forces with SHH in developing a distinctly liberal approach to behavioural public policy. The foil against which Oliver develops his approach is what he describes as the soft paternalist approach expounded by thinkers such as Thaler and Sunstein (2003, 2008). They seek to use behavioural insights to advance policies that centre on establishing a ‘choice architecture’—a way of framing the options between which decisions will be made, etc.—that encourages people to act in line with their true, underlying preferences. Thaler and Sunstein’s ‘libertarian paternalist’ approach supposes both that people have an underlying set of ‘true’ (stable, consistent, and context-independent) preferences but also that—because of the influence of various behavioural biases—they fail to satisfy those preferences as effectively as they might. Policy then centres on establishing a ‘choice architecture’—a way of framing the options between which decisions will be made, etc.—that encourages people to act in line with those true, underlying preferences. Oliver objects to this approach for two reasons. First, like SHH and others, he argues that experimental evidence indicates that “there is no underlying fixed and stable preference standard” upon which nudge-oriented policymakers can gain sufficient purchase for their preferred policies to be successful. Second, and also like SHH, Oliver argues that such an approach to policy fails to do justice to the importance of people’s autonomy. Soft or libertarian paternalism “is an undesirable policy direction if one wishes to preserve individual autonomy”, Oliver argues, elaborating as follows:

Nudging populations towards particular preset outcomes conflicts with the fact that people have various and varied desires that they might wish to fulfil in their own lives. It may well be that outcomes-based welfare is one thing that most people to some considerable degree desire, and for some people it may be all that they desire. But for most people, that which offers welfare in this sense differs across and (in different contexts and at different times) within people. We may, for example,

---

2 For another important and influential argument to the effect that, as Oliver puts it, “there is no philosophical or psychological basis to believe that an inner rational agent exists”, see Sugden (2018).
willingly sacrifice our own personal outcomes-based welfare in order to keep a promise, to limit the harms that we might otherwise place on others, to satisfy our own conception of fairness, to maintain a level of dignity that we believe is fitting to our position, or a host of other factors. Priorities differ across people; indeed, some people may desire to have no priorities at all, which, from a liberal perspective, is perfectly acceptable.

For Oliver, therefore, as for SHH, policy should rely on autonomy, not on preference-satisfaction, as its main normative standard.

In making the liberal case for autonomy as a normative benchmark, Oliver adopts a similar position to SHH in arguing for a ‘constitutional approach to behavioural public policy’ according to which “attention is shifted away from the focus on outcomes that is typical in the paternalistic frameworks and on to the rules that constrain and enable actions.” As SHH puts it,

public policy should be concerned with the character of the rules that constrain and enable behaviour (e.g., how procedurally fair they are and how much freedom they allow) and not simply the behavioural outcomes themselves (e.g., the extent of preference satisfaction). (Hargreaves Heap 2017: 253.)

This is a procedural approach, whereby policies are understood as changes in the ‘rules of the game’ and are evaluated by reference to their contribution to people’s autonomy rather than according to whether they further specific behavioural outcomes. Thanks to its emphasis on developing a framework of rules which affords people considerable freedom, rather than nudging people into satisfying the preferences of some chimeraical inner rational agent, this is—in both Oliver’s and SHH’s opinion—a liberal approach to policy. Oliver argues, moreover, that it is reminiscent of the form of liberalism advocated by the great American journalist and author Walter Lippmann in his influential 1937 book, *The Great Society* (Lippmann 1937). In that book, Lippmann sought to revive liberalism in the face of the rise of collectivism in the 1930s, arguing that “great and salutary changes in human relations can be and usually have been effected not by commands from on high but by amending the laws under which men deal with one another” (Lippmann, quoted by Oliver).³ Drawing on SHH and Lippmann, as well on his own recent book on The Political Economy of Behavioural Public Policy, Oliver argues that “the future of behavioural public policy would be best moulded upon an anti-paternalistic liberal vision – a vision that eschews a laissez faire ethos and yet respects the right for people to pursue their own desires in life so long as they do not impose substantive harms on others” (also see Oliver 2023).

As suggested by the title of his paper—namely, “What to do About Preference Change? Lessons from ‘Philosophy, Politics and Economics’”—Shaun Hargreaves Heap develops his earlier work exploring the significance of behavioural evidence about the context-dependence of people’s preferences. For SHH, preference change is a “blind spot” for economics in the sense that “it has been ignored ... without good reason, and with significant consequent damage for the discipline.” SHH takes as his benchmark the Arrow-Debreu model of general equilibrium that achieved considerable prominence in economics from the 1950s through to the 1970s and beyond. He argues that much important work in economics in the 1970s, 1980s and 1990s involved the relaxation of assumptions made by Arrow and Debreu about the exogeneity of particular economic factors (as, for example, when Romer’s work on endogenous

³ For more on Lippmann, see Steel (1980) and Goodwin (2014). SHH also draws on Lippmann’s work (see Hargreaves Heap and Lewis 2023).
growth theory sought to relax the assumption of exogenously-given technology). The principal exception to this pattern of seeking to ‘endogenise’ various factors taken as given by Arrow-Debreu was people’s preferences, which in SHH’s history of the discipline over the past half century have remained steadfastly given or exogenous. SHH illustrates how this reluctance to analyse preference matters for explanations in economics by arguing that accounts of two of the most important macroeconomic developments in rich countries over the past few decades, namely declining productivity growth and rising inequality, might have benefitted by taking preference change more seriously.

SHH then builds on his earlier work on the implications of preference change for welfare economics (SHH 2013, 2017, 2023), arguing that preference-satisfaction ought to be replaced as the benchmark for normative evaluations by a procedural or constitutional alternative that concentrates on the desirability of the rules or procedures under which people make decisions (rather than on the consequences of those decisions per se). For SHH, following John Stuart Mill, the key attribute upon which such evaluations should focus is the extent to which the rules in question “encode the principle of liberty.” Understood in its ‘negative’ sense (Berlin [1958] 1969), liberty is important because it enables people to “engage in experiments in living and to reflect ... on those experiences and so come to have reasons for action. This is the process through which we gain a sense of being autonomous, of being the author of one’s actions.” As noted earlier in this introduction, autonomy is a key value for SHH and he goes on to argue—invoking a more ‘positive’ notion of liberty—that a concern for autonomy, rather than preference satisfaction, implies that the rules governing people’s conduct should ensure that all people possess both the resources and the capabilities, in Sen’s sense of that term, required to engage in, and reflect on, experiments in living. It is also important to note that for SHH, again following Mill, liberty in the negative sense is not a license to do anything but is instead constrained by the ‘harm principle’ (that is, by the requirement that a person’s actions do not impose harms on others). What counts as a harm—or, put slightly differently, what rights people have to not be harmed—is determined by the rules prevailing in society and those, in turn—like the precise capabilities people are thought to need—are determined through the political process. And this is how, to return to the title of SHH’s paper, a PPE (philosophy, politics and economics) or political economy approach becomes significant for answering the question of what to do about preference change in economics; the need for a political process to answer such questions implies that a consideration of “politics is necessary for the consideration of what arrangements for collective decision making of this kind will render legitimate the concrete decisions over what capabilities and what counts as a harm and how to compensate for them” (also see Hargreaves Heap 2020).

References


