makers generally dampen crowding-out policy effects, or even trigger crowding in, by publicly reinforcing a positive interpretation? Or does any such interaction effect in turn depend on policy details that render such a message more or less credible?

Other questions about policy-communication interactions are likely to be new versions of old problems long examined in social science. For example, when is the information flow intense enough to trigger any sort of policy design effect, one way or the other? And to what extent is the information coded with partisan and ideological cues such that only those positively predisposed to the sender of the message will agree and be affected by policy design? Alternatively, perhaps all major political parties (or, say, both parents in a family) propagate the same consensual frame. Returning to the household chore example, perhaps both the author and the mother united around the same (cynical or social) message about the meaning of incentives? In such cases, policy design should matter more than usual and in the same direction regardless of predispositions towards individual parents or parties (in whatever direction reinforced around the kitchen table or in the public sphere). Of course, consensus in the interpretation of policy is probably unusual in family life, democratic politics, and other situations.

Staffan Kumlin
University of Oslo, Department of Political Science
staffan.kumlin@stv.uio.no

References

Kumlin, S. and I. Stadelmann-Steffen (eds). 2014. *How Welfare States Shape the Democratic Public: Policy Feedback, Participation, Voting, and Attitudes*. Cheltenham, UK: Edward Elgar Publishing.

Soss, J. and S. Schram. 2007. 'A Public Transformed? Welfare Reform as Policy Feedback.' *American Political Science Review* 101 (1): 111–127.

Prosociality Matters, But Let's Not Throw Out the Knaves Assumption for Heterogeneous Populations

What are the conditions for the emergence of cooperation, solidarity, and a 'social order' in groups, communities, and societies? Following Machiavelli, Hobbes, and, in a sense, the British Moralist tradition (Hume, A. Smith), many thinkers have answered this question by pointing out the role of appropriate institutions, which provide constraints and incentives (material rewards or punishments) to motivate individuals who are dominantly self-interested into contributing to public goods. Using the *homo economicus* model of man, that is, (complete) rationality and self-regarding preferences, it seemed safe in this tradition to argue that cooperation in social dilemma situations requires external incentives.

In the social sciences, classical sociologists like Durkheim and Parsons argued against this rationalist-utilitarian tradition. Durkheim objected that modern societies could not achieve cooperation via voluntary contracts among rational and self-interested agents because contracts are binding only owing to pre-contractual elements that are provided by 'society', that is to say by legal institutions, and, most importantly, the internal sanctions of morality. In a similar vein, Parsons developed what has been called the 'normative' solution to the Hobbesian problem of social order. Cooperation in society depends on categorical commitments to specific norms of cooperation that require that agents act, from time to time, against their narrow self-interests. Parsons argued against one fundamental premise of economics that he called the assumption of a 'randomness of ends'. Parsons believed that without any restrictions with regard to individual preferences (preference neutrality) a stable social order cannot be achieved. In recent decades, many economists have explicitly or tacitly adopted many ideas from sociology and

psychology to prepare a paradigm shift in the economic model of man. Representative agents are no longer modelled exclusively as rational egoists or *homines economici* but as boundedly rational individuals with social preferences of various kinds.

In this book, Samuel Bowles explores the role of social preferences in the emergence of the institutions of a good society. Since parts of the book were given as lectures at Yale University in 2010, the book's style is quite informal and comprehensible even to readers without much background in economics and game theory. However, this sometimes comes at the cost of some inaccuracy and lack of clarity. The book focuses on the impact of human motivations on social institutions, broadly conceived, in a twofold sense. First, Bowles refers to results from empirical research in behavioural game theory, in particular laboratory and field experiments, which demonstrate, as argued by the author, the relevance of social preferences for the emergence of cooperation, fairness, or altruistic behaviour. The experiments that are covered represent some of the standard work horses of experimental games (Prisoner's Dilemmas, Public Goods, Ultimatums) and, in addition, some experiments on interactions in natural ('field') situations. Second, the book refers to problems in normative social theory: What kinds of behavioural or motivational assumptions are appropriate if 'legislators' want to design good social institutions or constitutions? Bowles quotes (as the book's motto) Hume's famous advice to follow the maxim 'that every man must be supposed a knave: Though at the same time, it appears somewhat strange, that a maxim should be true in *politics*, which is false in *fact*.' Notice that Hume as well as Adam Smith both argued that humans sometimes and under certain conditions empirically act in accordance with a principle of 'sympathy', that is, they are able and willing to take the role of their interaction partners and identify with their respective interests. Nevertheless, legislators who try to construct efficient social institutions should not assume that such pro-social motives prevail. In contemporary constitutional economics Buchanan and Brennan endorsed Hume's maxim with regard to the design of basic societal institutions: '*Homo economicus,* the rational, self-oriented maximizer of contemporary economic theory, is, we believe, the appropriate model of human behavior for use in evaluating the workings of different institutional orders' [Brennan and Buchanan 1985: 61]. The rationale for this normative maxim is not that human behaviour is empirically governed by rational egoism under all circumstances but that even in populations with a majority of actors who are endowed with pro-social motivations these motives may be driven out: 'the narrow pursuit of self-interest by a subset will induce all persons to behave similarly, simply in order to protect themselves against members of the subset' [ibid.: 68]. In other words, if institutional rules are designed on the premise of the pro-social behaviour of the target actors (who are supposed to follow the rules), there is a risk of triggering a crowding-out process with respect to these motivations ('Gresham's law of politics').

In this book, Bowles defends, so to speak, the opposite position that using what might be called the Hume-Buchanan maxim would neither be possible nor necessary to design good institutions. Bowles not only extensively cites and describes empirical results ('hard evidence') about social preferences but in addition gives an interesting, albeit short, discussion of the 'mechanism design' programme in economic theory. Mechanism design theory (as, for example, developed in auction theory) rests on three adequacy conditions: preference neutrality, voluntary participation, and Pareto efficiency. Bowles (cf. pp. 168–174) alludes to an impossibility result, which he calls the 'liberal trilemma',

namely that the three goals cannot be realised simultaneously. In other words, it is impossible to construct rules for a constitution of knaves that work (p. 174). On the other hand, social preferences are, according to Bowles, quite widespread and effective in certain populations.

The persuasiveness of Bowles's central theses to a large degree depends on what is meant by 'social preferences' and on the supporting empirical evidence. In fact, Bowles defines social preferences very broadly. They do include preferences that depend on payoffs, perceived intentions, or other aspects of elements of the members of a reference group (or interaction partners in an experimental game). In experimental game theory there exist standard models that represent inequity aversion or fairness preferences by utility functions with (mathematical) arguments which depend on a comparison of the ego's payoffs with others' payoffs. (By the way, some of these models translate familiar ideas from social psychology and sociology into the language of utility functions—for example, relative deprivation and relative gratification.) Bowles's treatment of social preferences lacks conceptual clarity when he states (cf. p. 180) that in a repeated Prisoner's Dilemma there may evolve a 'social preference' to conditionally cooperate if the shadow of the future is large. The standard approach to repeated games treats conditional cooperation as a *strategy* in the repeated game—not as a preference. It seems to me that it would be conceptually disastrous to break down the demarcation line between preferences and strategies (i.e. elements of the opportunity set) at this point.

Furthermore, and most importantly, Bowles treats intrinsic motivations as elements of the set of social preferences. It has been emphasised before by Bruno Frey and other economists who adopted ideas from psychology that humans may be motivated to perform certain actions because these actions are valuable per se. Under certain conditions intrinsic preferences (for example, to cooperate) will be crowded out if external interventions provide material rewards or punishments. There is a comprehensive set of anecdotal and experimental evidence from psychology and also, but less so, from experimental economics that demonstrates such crowding effects. Bowles argues that in cases of intrinsic preferences a central assumption of neoclassical economics is violated, namely additive separability. In the context of intrinsic preferences this means that externally provided incentives and intrinsic motives to perform an action do not work additively but there may be interferences such that external incentives on the one hand *increase* the tendency to choose the relevant action and on the other hand *reduce* the strength of intrinsic motives. The *total effect* may be such that the propensity to perform the desired action is reduced due to the incentives—contrary to the goals that the 'social planner' wants to accomplish. As a case in point, consider the effects of intensive external monitoring or variable rewards (piece-rate payments) on workers who, prior to the intervention, were intrinsically motivated to perform their tasks. In such cases, it may well be that interventions decrease the quality and quantity of workers' outputs because they undermine intrinsic motivations.

Intrinsic preferences are situation-dependent in the sense that changes in the external environment or a new 'framing' induce actors to consider a different subset in their repertoire of utility arguments, they do not necessarily yield long-term endogenous preference changes (see p. 117). From this perspective it follows that institutional changes (via external interventions) should be undertaken with great care in order to avoid effects that oppose the legislator's intentions. In the final chapter of his book Bowles presents advice to legislators who want to cope with the sep-

arability problem. The list of advice covers several types of situations including those cases where intrinsic preferences are presumably absent in the population. In this case, the advice is to adopt the standard theory (pp. 205–207).

Bowles's book is highly readable and very inspiring. It provides the reader with an original informal interpretation of social preferences (including intrinsic motives) and describes a carefully chosen set of experimental studies which are re-analysed or discussed in some detail in order to support the book's main argument. However, some readers, in particular those with experience in conducting behavioural experiments, will object that many experimental results (e.g. with regard to the effects of punishments in public good situations) are not very robust or culture-sensitive. Replications often significantly fail or demonstrate effects which are much weaker than some prominent results from the literature. In many cases there are multiple alternative interpretations of experimental results, some of which would be in line with more standard ('homo economicus') approaches. This is due to the fact that in most experimental games there are multiple Nash equilibria, in particular for populations with pro-social motives.

With regard to normative constitutional theory, Buchanan's approach towards the construction of universal rules which are valid for an extended time horizon and possibly for a population with a high degree of diversity seems to me still warranted. Since such constitutional choices will necessarily be made under conditions of high uncertainty with regard to the properties of the affected agents and the specific distribution of social preferences, it may be a 'quasi-risk aversive' choice (Brennan and Buchanan) to use the standard homo economicus assumptions as a first approximation. For the time being it seems to me that Bowles's approach will be most effective in cases of small and stable social situations with a fixed population of interacting individuals whose characteristics are well-known and homogeneous. In populations with heterogeneous individuals from different cultural contexts and with considerable migration, 'legislators' will in general not be able to gather much ex-ante information with regard to the specific mixture of preferences. This information, however, is necessary to estimate whether or not pro-social motives will be undermined to a significant degree by incentives that are provided by the constitutional rules.

Thomas Voss
University of Leipzig
voss@rz.uni-leipzig.de

References

Brennan, G. and J. M. Buchanan. 1985. *The Reason of Rules—Constitutional Political Economy*. Cambridge: Cambridge University Press.

Institutions for 'Knaves' May Still Backfire—Liberal Civic Culture Needs to Foster Social Preferences: The 'Institutional Gardening' of Citizens' Virtues

Samuel Bowles, Professor Emeritus at University of Massachusetts and now Director of the Behavioral Sciences Program at the Santa Fe Institute, New Mexico, is an economist with a long-standing interest in institutions' interactions with individual preferences and actions. Much of his work is dedicated to challenging deeply entrenched assumptions of economic theories about free markets and people's choices. This topic is also at the centre of the book, which resumes decades of work and thinking on the question of how better policies can be designed by taking account of the full range of motivations driving people's actions. For a long time, the standard economist's tale of why people do what they do has merely focused on costs and benefits