

Hobbesian and Contractarian Constitutions

Geoffrey Brennan and Giuseppe Eusepi

1 Contractarian Methods and Constitutional Analysis

Within the constitutional political economy tradition, the predominant normative framework is contractarian. That is, the standard test in evaluating political institutions is whether such institutions could conceivably [or more rarely, actually did] emerge from free agreement among rational individuals. Clearly, the application of this test will depend on a range of details: whether the agreement in question is actual or ‘conceptual’; what exactly it means to say that such agreement is “free”; what rationality requires and whether the requirement of rationality seriously restricts the possible institutional arrangements that might emerge from actual or hypothetical agreement; and whether the contract in question must be between individuals or might be between groups of some kind. However these issues are resolved, the central role of agreement and its analogical connection to market exchange are central elements in standard “constitutional political economy” (CPE henceforth).

This strong contractarian flavour reflects the influence of Buchanan on the CPE enterprise.

G. Brennan (✉)

School of Philosophy, ANU College of Arts and Social Sciences,
The Australian National University, Canberra ACT 2601, Australia
e-mail: Geoffrey.Brennan@anu.edu.au

G. Brennan

UNC-Chapel Hill, Chapel Hill, USA

G. Brennan

Duke University, Durham, USA

G. Eusepi

Faculty of Economics, Department of Law and Economics of Productive Activities,
Sapienza University of Rome, Rome, Italy
e-mail: Giuseppe.Eusepi@uniroma1.it

And many of the positions taken in CPE analysis correspondingly reflect Buchanan's views on the "details" already enumerated. So, for example, Buchanan is determinedly individualistic; and reckons 'rationality' to be a distinctive property of individuals—and only contingently a property of groups, for well-known prisoner's dilemma reasons. Accordingly, only agreements among *individuals* can have any ultimate normative authority. And Buchanan is inclined to understand "rationality" 'thinly', in the manner characteristic of ordinary economics. That is, rationality is understood to require that agents act to promote their ends—but the ends in question can be [almost] anything at all, provided they are not blatantly self-contradictory. In addition, the reference to market exchange as a benchmark notion in understanding the notion of agreement reflects a particular conception of agreement—a conception that arises naturally among economists but is less common among political and moral philosophers. In short, the 'contractarian approach' followed in most 'constitutional political economy' is a distinctive one and differs in a number of ways from other kinds of contractarianism—Rawls's, Gauthier's and Scanlon's for example—that inhabit the political and moral philosophy literature.

Moreover, it is clear that normative constitutional analysis might be pursued from other, non-contractarian, normative positions. There seems to be no particular reason why contractarians as such should have a monopoly on concern with institutional arrangements. Alternative sets of political economic 'rules' might be evaluated from utilitarian or neo-republican or libertarian or egalitarian positions; and indeed political theory abounds with such exercises. Constitutionalism does not necessarily involve contractarianism.

There is, however, a suggestion in the Buchanan formulation that the obverse logical relation might apply—that is, that contractarianism implies constitutionalism. Much of the Buchanan interest in constitutional rules arises through an apparent conviction that a shift to the constitutional level of analysis is necessary if contractarian modes of thinking are to have any substantive content. In *The Calculus of Consent*,¹ for example, the explicit reason for shifting to the analysis of decision rules is that the contractarian ideal of unanimity is infeasible [or at least 'too costly'] at the in-period level. The intimate connection between unanimity and the Pareto criterion of the "new welfare economics" is a recurring theme in Buchanan's writings.² So is the notion that informational limitations imply that actual exchange (or actual agreement) is the only authoritative test of mutual benefit. Consider:

¹Buchanan and Tullock (1962).

²See, for example, "The Relevance of Pareto Optimality"; "Positive Economics, Welfare Economics and Political Economy" in Buchanan (1999).

... Unless the observing economist is assumed to be omniscient, his classification of a final position as non-optimal can never be more than a conjectural hypothesis that is impossible to test. If members of the group do not explicitly choose among final positions in the appropriately defined welfare space, the hypothesis that some members of the group can be better off by a change remains empty. ... By contrast, the classification of an organisational rule as non-optimal can be considered as a hypothesis that is subject to conceptual testing. If a presumed or apparent non-optimal rule cannot be changed through agreement among members of the group, the hypothesis stating that the rule is non-optimal is effectively refuted.³

The idea here seems to be that by moving to the level of choice among rules the scope for agreement is enhanced, and this is both because the informational requirements are less extreme at the constitutional level and because it is not necessary that ‘optimal rules’ always produce ‘optimal results’. What is clear, however, is that if there should be a set of political arrangements which would leave all citizens better off than in the status quo—a set, that is, such that a move from the status quo arrangements to the alternative would command agreement among all [rational] citizens and can indeed be shown to do so—then these political arrangements should be implemented.

It is worth noting that the whole Buchanan approach is framed against several feasibility considerations. One of these is access to relevant information; hence, the reference to the economist’s putative “omniscience”. Here Buchanan is gesturing towards the abandonment of utilitarian methods under the onslaught of Robbins and Hayek—a gesture that sits well with his own expressed sympathies for subjectivism.⁴ Actual exchange is capable of revealing the values that parties place on options available; in the absence of such exchange, the economist has no privileged access to information about values and indeed it is not entirely clear whether such value can be said to exist in the absence of agents’ choice behaviour.⁵ Another ‘feasibility’ consideration that has played a critical role in the development of public choice theory specifically concerns assumptions about human motivations—particularly in relation to agents in their political roles. The “benevolent despot” construction has been a focus for public choice critique from the very founding moments of the discipline. The idea that policy advisors and/or their politician/bureaucrat advisees can be appropriately modelled as motivated exclusively to pursue the public interest (however, exactly discerned) has struck a succession of public choice theorists as hopelessly “romantic”—hence, the description of public choice theory as “politics without romance”. This assumption, no less than the idea that political considerations are irrelevant in policy choice—the ‘despot’ aspect of the ‘benevolent despot’—is something that public choice expressly aims to set aside. If we focus on agents as they are, if we treat human imperfection as a basic constraint, then we must design institutions that accommodate such imperfection as best we can.

³Buchanan (1999), p. 228.

⁴See in particular his *Cost and Choice* (1969) and “The Domain of Subjective economics” (1982).

⁵See Buchanan (1991).

In this motivational respect, constitutional choice also offers some relief. In a way similar to that under Rawls' veil of ignorance argument, natural selfishness gives way to a measure of (rational) general interest as the position of each becomes increasingly uncertain. The argument here is familiar and does not require repeating. But it is worth noting that this consideration expands the domain of the feasible and provides us with an independent 'reason for rules'. And this reason for rules, because it involves increasing the likelihood of agreement over the objects of choice, is specifically *contractarian* in character.

2 The Hobbesian Element in Constitutional Political Economy

An obvious benchmark case in discussing 'the reason for rules' and the choice among alternative possible sets of rules, is the case of 'no rules'. If it is to be shown that agents actually *choose* constraints to be subject to, over and above those that are imposed by forces of nature, it needs to be shown why choosing such "facultative" constraints⁶ is rational. In analysing the 'no rules' scenario, Buchanan and the CPE tradition have looked primarily to Hobbes. Buchanan is no libertarian anarchist. He takes the necessity of government seriously; and indeed, the construction of government out of the rational agreement of ordinary agents is seen to be a paradigmatic instance of constitutional exchange.⁷ The attractions of Hobbes in this connection are self-evident. Hobbes is, on most readings, as close to a *homo economicus* political theorist as it was probably possible to come in the seventeenth century. Hobbes, like Hume, nurtured the ambition—largely shared by modern economics—of developing a theory of human behaviour on all fours with physical science, and with all the rigour and elegance of Euclidean geometry. And Hobbes, exactly like Buchanan, sought to derive an account of the legitimacy of government—and of the exercise of coercive power by government—from an argument based in consent of the governed. But Hobbes, unlike Buchanan, does not develop that argument directly. In the Hobbesian conception, it is an error to impose a requirement that government be directly constrained by popular will. For Hobbes, although citizens plausibly choose to be subject to governmental rule, they cannot choose to be subject to specific laws and policies of their own choosing. The contractarian element in Hobbes plays itself out at some remove from day-to-day political process. And this fact makes the Hobbesian account very different from the standard contractarian one. Indeed, it is not at all clear that Hobbes represents a particularly congenial starting point for the CPE tradition. If one takes Hobbesian anarchy as a point of departure seriously, the subsequent analysis may have to develop along lines rather different from those familiar in CPE discourse.

⁶See "The Relevance of Pareto Optimality" p. 211.

⁷As argued, most notably, in Buchanan's *The Limits of Liberty*.

Let us elaborate briefly. In one early version of the contractarian tradition, associated with Wicksell (1896) and Lindahl (1919) from which much of Buchanan's contractarianism derives, the laws and policies to which people are to be subject are construed as "chosen" directly by the citizenry. Politics is thought of as a direct analogue to *n*-person exchange. But as Buchanan and Tullock (1962) emphasise, this can only be strictly true if the collective decisions about laws and policies are subject to unanimous consent. And unanimous consent is self-defeating in the sense that a decision rule for collective decisions of unanimity would itself be unanimously rejected. Citizens would foresee that under any system of political decision-making in which each individual citizen holds a veto, virtually nothing would ever get done. Citizens would, so Buchanan and Tullock argue, opt for a less-restrictive decision rule. Under a 66 % majority requirement, for example, citizens might reasonably expect that they would be members of the decisive majority two-thirds of the time; and hence that the benefits enjoyed when in the majority would exceed the expected losses to be made in the one-third of cases when they are in the minority. Being "coerced" into paying for collectively provided projects from which I derive no benefits becomes the price I am prepared to pay in order to "coerce" others to contribute to the collective projects from which I do benefit.⁸

In this conception of politics, there is no 'principal/agent' issue. A 66 % majority rule is supposed to act as a perfect agent for whatever majority of individual citizens happens to constitute the two-thirds coalition. But at least in representative political systems (which represent the overwhelmingly predominant practice), individuals do not vote for policies issue by issue—or even for rival packages of policies: they vote for agents (individual political candidates or political parties) to whom they delegate substantial powers of policy determination. And those candidates will have interests of their own which need not align in all cases with those of the citizen voters—and in some matters can be expected to diverge systematically.

Hobbes's thought involves carrying that case to the limit. The individual or group to whom the power to govern is delegated is treated as having total discretion to act as he/she/they see fit. Individuals are offered a choice concerning whether they are to be 'in-laws' or 'outlaws' but they are not offered a say in what the 'laws' will be—the decision about the content of laws is exercised exclusively by the governing power.

There is one interesting 'half-way-house' in the CPE tradition. This is the case laid out in Brennan and Buchanan's (1980) analysis of the power to tax. In the model of government deployed in that treatment, governments are taken to be

⁸We have placed "coercion" in inverted commas here because there is a question as to whether what is at stake here is 'coercion' at all. It is after all something that by hypothesis all consent to. Perhaps, it can be thought of as the collective equivalent of being forced to pay for the car that one drives off the car lot, or the food one buys in the supermarket. For, note that the (legal) requirement to pay for such things reflects a collectively defined and enforced property order to which the purchaser may not have explicitly consented and perhaps in some (few) cases may not on balance benefit from.

totally unconstrained by in-period electoral processes. Governments simply do whatever maximises the government's (that is, the political agents') preferences. However, in the Brennan/Buchanan model, citizens do have some residual constraining power: they have access to a possible 'constitutional restriction' on the *extent* of the taxing power. This model is a move in the direction of the Hobbesian approach—in that citizens have negligible influence on what governments do. But it falls well short of the full Hobbesian thought experiment because the Brennan/Buchanan "Leviathan" is subject to a broad fiscal constraint that specifies the extent of the 'power to tax'. Hobbes' Leviathan is restricted to constraints (if any) that it is somehow profitable (and feasible) for a rational Leviathan to impose on itself. If the set of constraints that fit this requirement is non-empty, that set will constitute the "Hobbesian constitution". And it is, we hope, perfectly clear that any such set is likely to fall well short of what is envisaged to emerge from the 'constitutional agreement' envisaged in Buchanan and Tullock (1962).

Put another way, constitutional political economy has plausibly constructed a 'reason for rules'. But it has operated as if rules of the kinds over which choice is presumed to be exercised are actually feasible. That is, only if we can take it that we *can* collectively impose on ourselves rules of a kind that we can reasonably trust each to abide by, we can reasonably enquire as to which set of rules is best. If there are no rules that agents will abide by, then although we may all in principle benefit from the imposition of such rules, there is little point in pursuing analysis because rules of this kind are simply infeasible. The whole exercise becomes one in futility. If we are to take the 'no rules' benchmark not just as an imaginative construct but as a plausible social equilibrium—as both Hobbes and Buchanan do—then we have to explain how it is exactly that escape from that equilibrium is possible. How is it possible that 'rules' do the work they are purported to do?

Perhaps one might respond to this challenge with the observation that as far as one can see, people do seem to obey rules in enough numbers to make the assumption of rule-following behaviour plausible, and so the question of the feasibility of rules is itself a rather specious one. But if this is so, then one does not seem to have any account of why government is needed for the enforcement of the chosen rules; or indeed why the whole exercise of justifying the existence of government as an instance of 'the reason of rules' is not self-contradictory. Of course, this issue is a problem for Hobbes, no less than for CPE. And it is by no means clear what exactly Hobbes' own solution to it is; or whether that solution is coherent. However, it is of considerable interest to see how Hobbes tackles this puzzle; and what a plausible Hobbesian solution to it would imply about how the 'constitutional perspective' ought to be applied to specific issues of institutional design.

This is the aim of this paper. We begin with an attempt to construct a version of the Hobbesian argument that seems to us to be at least coherent. We then turn to a consideration of what this would mean for the 'constitutional perspective' and for a genuinely contractarian approach if the Hobbesian Constitution were the relevant one.

It perhaps goes without saying that the particular Hobbesian construction of the state of nature and the possible routes out of it is quite different from other scholars' construction of the 'no rules' case. If the contractarian position took John Locke (1689) or Jean-Jacques Rousseau (1762) as a point of departure, doubtless the challenges of reconciliation would be rather different. But as a matter of fact, it is Hobbes who figures in the CPE literature as defining the benchmark 'no-rules' case; and so it is Hobbes's logic with which CPE has to grapple.

3 The Hobbesian Story

A disclaimer is in order at the outset in this exercise. We are not historians of thought in anything except an utterly amateur sense. We do not claim that our reflections represent "what Hobbes really meant"; we do not see ourselves to be providing an interpretation of Hobbes in that sense. There are a number of extremely interesting (and highly varied) such interpretations. Our approach is to take what we see as various 'fixed points' in the Hobbesian account of emergence from the state of nature; and try to fit a minimalist coherent structure around those fixed points. It is in that sense that we tend to refer to the "Hobbesian logic" rather than to Hobbes *simpliciter*. Arguably, what we are engaged with here might be best described as *Buchanan's* Hobbes—Hobbes, as he figures in framing the broad constitutional contractarian story. For our main concern is with whether the Hobbesian logic as deployed in the CPE tradition, as we read it, is consistent with a standard contractarian story in terms of its implications for constitutional analysis.

There are good reasons for not getting committed to Hobbesian exegesis—quite apart from our own limitations. For example, there seem to be two different treatments of the central issues—one in *Leviathan* and one in *De Cive* (both published in 1651). Moreover, there are multiple "contracts" in play in these discussions, with different terminologies and apparently different concepts in the two treatments.

For example, in *Leviathan* we find two separate contracts:

- (i) *the pactum unionis*, through which individuals abandon "moral anarchy" and move to "moral order", in Buchanan's terms, and
- (ii) *the pactum subiectionis* through which all but the Leviathan are subdued to rules.

The *pactum unionis* involves individuals agreeing to treat each other as morally equal and to refrain from using the force to solve interpersonal conflicts. But the status of this particular 'contract' is unclear. In particular, while it is clear what individuals forego with the *pactum unionis*, it is much less clear what they actually get. One might for example think that morally equal individuals would feel equally bound by the contract they have agreed on. If that were so, however, the consequence would be that the state of nature is made up of self-constraining individuals.

Depending on the content of such self-restraining rules, life in the state of nature might be rather limited but hardly as nasty brutish or short as the familiar picture of Hobbesian anarchy is taken to be. Indeed, one might think that within a society of moral equals, *all* individuals, none of them excluded, would be equally entitled to take decisions. In this sense, the *pactum unionis* appears to be designed for some kind of *direct or pure democracy*. But in Hobbes's treatment, it seems to be deployed as simply the first step to *direct or pure absolutism*. It seems as if, for Hobbes, the *pactum unionis* was not to be viewed as a contract at all, but merely a sort of preliminary compromise between parties. From Hobbes' viewpoint, then, it is the *pactum subiectionis* that is the real contract with which individuals shift from the state of nature to the civil state. But this contract seems on its face to violate the moral equality stipulation of *pactum unionis* because by construction those who exercise the authority of the civil state lie outside its authority.

De Cive finesses the problems created by the dual contract, but the "patch", which Hobbes uses to do this, seems to involve a radical alteration of the whole theoretical framework. In effect, *De Cive* cancels the *pactum unionis*: there is only the *pactum de imperio*, and the precise relation of this latter contract to the categories used in *Leviathan* remains unclear.

Accordingly, we shall take a stick figure version of Hobbes—consisting of what we take to be the central features of the account (or what the standard interpretation seems to be in CPE circles)—and construct from that a basic Hobbesian logic.

The relevant fixed points are as follows:

- (a) the state of nature is nasty, brutish and short;
- (b) the state of nature is [at least] a short-run equilibrium;
- (c) exit from the state of nature is feasible only by virtue of establishing a sovereign;
- (d) the state under the sovereign is an equilibrium;
- (e) the state under the sovereign is better for all than the state of nature.

Note that any theoretical account that can satisfy the requirements of that story for a set of rational agents has several challenges. First, we have to show that the state of nature is a possible equilibrium in some meaningful sense. Second, we have to show that the state with sovereign is an equilibrium in that same sense. Third, we have to show that this latter equilibrium is the *only* equilibrium other than the state of nature. Fourth, we need to show that the sovereign equilibrium is better for all agents than the state of nature equilibrium. And finally, we have to show that the route from the one equilibrium to the other is accessible to agreement among the parties of some plausible kind.

The Buchanan strategy in this setting is to conceive of the state of nature as an n-person prisoners' dilemma. The 'equilibrium' outcome in this game is one in which each individual wages wars of pre-emptive aggression on all others—that is, the strategy of 'universal defection/non-cooperation' is indeed an equilibrium strategy for all rational players. We can also imagine that there might exist an agent who, if appointed to the role of enforcer of cooperative behaviour [or of agreements

by individual agents to engage in cooperative behaviour], both *could* and *would* enforce that behaviour.

But both the ‘could’ and the ‘would’ here are problematic. ‘Could’ first. How is it feasible for the sovereign to enforce the behaviour in question? What alchemy is wrought in the process of making that individual ‘sovereign’ that gives him the power to enforce that which he did not possess in the state of nature? Put another way, when the sovereign attempts to enforce his will, what makes me obey? And if the answer involves an appeal to the force of a promise I have made, why could I not make a promise to behave cooperatively [contingent on others doing so perhaps] directly with other individuals in the state of nature? If the sovereign can indeed enforce, why do we need him? That is one issue.

What of ‘would’? What is it that motivates the sovereign to do anything other than exploit the subjects who have (somehow) placed themselves under his thrall? How is it that agents have done anything other than replaced a state in which life is nasty, brutish and short for one that is nasty, slavish and all too long? Clearly, there must be an answer to this question, because otherwise it would not be rational for agents to leap out of the state of nature in any case other than that in which they stand a chance of being king. Or perhaps the latter possibility *is* the answer. That is, the expected return arises from agreeing to a regime in which someone is chosen at random to be king and the benefit on offer is just the benefit from enslaving all others. But if so, what binds the agent to adhere to the rules once the regime is in place? Once I discover that I am not the king, why should I rationally obey the one who is?

Presumably, Hobbes believed he had answers to these questions, although Hobbes scholarship seems somewhat at odds over what those answers were. In any event, we do not see ourselves restricted to those answers to which Hobbes might have had access. We are content to raid contemporary social science for plausible answers and let those scholars with expertise in Hobbes exegesis explain why it is, or is not, the case that these particular answers were those that Hobbes had in mind.

In the state of nature there is no external enforcement of promises. Hobbes makes clear that the only obligations that can be binding are self-binding ones. It is useful therefore to distinguish at the outset between two classes of promises: those that are self-binding; and those that require external enforcement. Denote these as type-A and type-B promises, respectively. Then it must be the case that the move out of the state of nature is in two steps. First, each makes a type-A promise, which is sufficient to establish the sovereign. However, to directly adopt the ‘cooperate’ strategy in the n-person prisoners’ dilemma must involve a type-B promise—one unenforceable in the absence of the sovereign. Otherwise, the state of nature could not be an equilibrium; and the external enforcement powers of Leviathan would not be required. Second, it must be the case that enforcement of cooperative behaviour under the sovereign does not require fulfilment of a type-B promise by any agent. These two requirements seem to constitute constraints on any coherent rendering of the Hobbesian story.

Consider an example of this structure that seems at least *prima facie* consistent with Hobbes’ own account. The characteristic case of a type-B promise is a promise

that would imperil one's life. Agents simply cannot be relied on to keep promises that involve them in risking their lives. A promise that involved the undertaking to risk one's life in the future would be "unnatural" and incredible. So, in particular, a promise to enter directly into a contractually supported common peace would be of this kind. There is simply nothing to ensure that I will, if I so promise, be exempt from attack because my promising does not ensure the promise being made or adhered to by others. Whenever anyone has reason to fear that he is under threat for his life, he will defend himself to the death, earlier promises notwithstanding. But lesser promises—say, to obey the sovereign's will whenever one's life is not at stake—are credible. So in particular, the promise to pay taxes and do 'public duty' of a non-life-threatening kind according to the sovereign's dictates can be taken to be (self-)binding. Now, suppose that external enforcement involves enrolling appropriately many against one, in such a way that, though the one defend himself to the death, the many are under no risk of death. Say that if the enforcers outnumber the non-cooperator by more than three to one, all of the enforcers are utterly safe. Then all can commit credibly to acting as enforcers of the sovereign's will in those numerical circumstances, though none can promise to abide by the sovereign's will unilaterally—or indeed to never resist attack should one be subject to such. Thus, even should the sovereign decide to execute one, for whatever reason, the one will resist; but the resistance will avail him nothing provided the sovereign has called on support in appropriate numbers.

If this reasoning provides us with grounds for thinking that creation of a sovereign is both necessary and sufficient for a leap out of Hobbesian anarchy, it does not provide us with reasons for believing that this leap will be better for all players. It does not then provide us with grounds for believing that the move will satisfy the contractarian test; and relatedly, does not necessarily explain why choosing to make the necessary type-A promise is rational for all agents. In order to provide such grounds, we need to say something about the rational conduct of the sovereign, once established.

Note first of all that the rationality test for ordinary agents only requires that the state under the sovereign be better for the agent *in an expected sense*. It may be expected that the sovereign will act in such a way as to make some agents worse off than they would be in the state of nature; and agents can know that this will be so for some agents. Provided only that the identity of such agents is unknown at the point where the type-A promise is made, then an expected return calculus is all that is required to make the giving of that promise rational. Whether or not an agent is subsequently made worse off does not set any logical bounds on what the afflicted agent can do. Once the sovereign is established, no single agent can unilaterally plunge society back into a state of nature.⁹ So for example there is no constraint that the leap out of anarchy must satisfy the Pareto test (i.e. make everyone better off) for the type-A promise to be rational. But we do have to show that agents have

⁹Buchanan seems to suppose the opposite in his "Before Public Choice"; this seems to us to be a departure from strict Hobbesian logic.

reason to expect that they *will* be better off under the sovereign on average. What grounds do they have for that belief?

Let us suppose that ‘support for the sovereign’ comes in degrees. In particular, within the terms of any feasibly enforceable law, there is a kind of principal–agent problem for the sovereign in ensuring that his subjects do what would actually be best for the sovereign. Suppose, in particular, that the sovereign has read his Smith and his Hayek. The sovereign knows in particular that the wealth of his nation will be maximised under a regime of only modestly modified *laissez-faire* with well-ordered markets. Of course, to maximise the wealth of the nation in the Smithian sense—that is, to maximise the aggregate income of his subjects—is of no direct interest to the sovereign. It only becomes ‘interesting’ to the sovereign to the extent that he can appropriate some portion of that wealth. And there will clearly be a trade-off between achieving maximal aggregate wealth and achieving the maximal amount for the sovereign. But, in particular, it seems possible that the best the sovereign will be able to do will be to levy a uniform tax rate on all income and otherwise allow trade and the division of labour to flourish unabated. In other words, the sovereign will rationally implement a regime of ‘natural liberty’ in Smith’s sense, supplemented with a revenue-maximising fiscal regime. And that revenue-maximising fiscal regime will interfere minimally with the functioning of the economic order because that is the way to maximise fiscal returns. Of course, the sovereign could attempt to manage the activities of his subjects in the manner of a slave–owner managing a community of slaves—that is, specifying the tasks of each and enforcing punishments for any failure to reach specified targets. But the sovereign’s familiarity with Hayek reminds him that such ‘managed economies’ work rather poorly, and that the extensive division of labour characteristic of truly ‘wealthy’ economies can emerge only when each individual agent is free to respond to market signals and is provided with relevant incentives to do so. It will be useful to have a term for this regime. Let us call it the ‘modified system of natural liberty’—MSNL. The characteristic features of MSNL are as follows: each agent has effective property rights in the fruits of her own labour, apart from the share which the sovereign appropriates in taxes; and that there are all the institutional facilities of the free market, including civil protection for all fiscally relevant rights, rules for the free exchange of rights and/or the fruits from them, and enforcement of contracts concerning such exchanges. The ‘rights’ of all parties against other private parties will be enforced up to net revenue-maximising levels. And given the residual interest that the sovereign has in the economic flourishing of each and every citizen and the economic flourishing of the whole, it seems plausible that the sovereign will have considerable incentive to enforce the rights of citizens against violations by other citizens with considerable vigour.

More perhaps needs to be said about the precise properties of the revenue-maximising tax regime; but if we think of it as involving a simple uniform-rate income tax, set at an appropriately high level [well short of 100 %], this approximation will do for the purposes of the argument. Effectively what the rational sovereign is doing is to devise a structure of private rewards for citizens such that the share left with them is sufficient to induce them to produce for him the

maximum revenue of which they are capable. This is an idea that is familiar from the recent economic literature on rational dictatorships—though of course, for real-world dictatorships, the threat of rebellion creates constraints of which Leviathan in the strict Hobbesian model is exempt by construction. If the sovereign is genuinely secure—if, that is, the type-A promises that agents made in the state of nature are truly binding on citizens—then the sovereign will have no reason to depart from his MSNL regime. That is the regime that it is rational for the sovereign to pursue—or rational, at least, for that sovereign who is motivated by the desire to maximise his economic power, the splendour of his court and his own private fortune. If citizens know all this, then they have reason to trust that their lives will go much better for them on average than in the state of nature. A simple reliance on the sovereign's prudence will give them adequate reassurance of a better future under the sovereign than under a 'no-rules' regime.

However, it is to be emphasised that the sovereign here is not bound by any contract. The sovereign provides no undertakings to citizens in this picture and is not bound to provide any undertakings. The only contractual element is the set of type-A promises made by citizens in extricating themselves from the state of nature to abide by the will of the sovereign. The sovereign himself remains in the state of nature with respect to any undertakings made, namely, that they must be 'self-enforcing'. And so far, no undertakings of any sort by the sovereign have entered the picture. The idea that the sovereign enters into any contract with citizens in the establishment of the Leviathan order has no place so far in the logic of the argument. So, to this point, we have no Hobbesian Constitution: simply predictably rational action on the part of the sovereign, once established.

4 The Hobbesian Contract

A Hobbesian Constitution, if one is to exist, must depend on the sovereign's having reasons to bind himself in certain ways. That is, it must be the case that strictly rational action at every point does not involve things going as well for the sovereign overall as they might if the sovereign were to resile from [the possibility of] rational action in certain contexts. We know that Hobbes thought that this situation could arise. Citizens in the state of nature will do things [rationally] that they would do better to forswear; citizens *can* bind themselves [if only by type-A promises] and it is rational for them to do so. Since the sovereign remains in the state of nature in the sense that there is no external enforcement, the sovereign has access to type-A promises, which are credibly binding, if he should wish to avail himself of them. Why would he?

Note that the regime of the revenue-maximising tax depends on a measure of trust on the part of individual citizens. For the proper functioning of the MSNL, at least some agents will have to accumulate capital. And they will, to some extent under the uniform income tax regime, reveal information about their income-earning capacities that it would be logically possible for Leviathan to

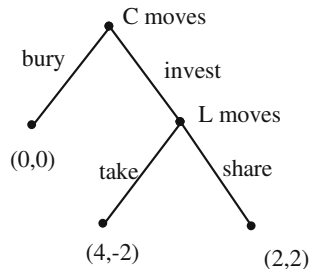
exploit. That is, Leviathan might be tempted to levy higher tax rates on those who reveal an exceptional income-earning capacity and/or to confiscate the accumulated capital of those whose accumulations reach significant magnitude. Unless the citizens can be sure that Leviathan will not give into such temptations, the possibility will moderate the income-earning/accumulation behaviour of citizens in a way that is income-reducing for Leviathan overall.

Consider for example the single-shot trust game in this setting. Here, citizen [C] gets to choose first whether to accumulate, thereby rendering her liable to expropriation in the next period. If she chooses not to accumulate, she and Leviathan both receive a return of zero. In the second period, Leviathan [L] gets to choose between simply applying the income tax on the return—in which case both C and L receive a pay-off of 2—or to apply a confiscatory wealth tax in which case L gets 4 and C loses 2. Thus, the pay-off structure is depicted in Fig. 1—which is the standard ‘trust’ game.

The characteristic feature is that there exists an equilibrium in this game—namely the no accumulation outcome—which is Pareto-dominated by another technically feasible outcome—the ‘income tax’ outcome. However, the income tax outcome is, in this one-shot game, ruled out by Leviathan’s rationality. If L chooses the higher pay-off open to him at any choice node, L will choose to expropriate, leaving C worse off than if she had not accumulated. Ergo, C will not rationally accumulate and both C and L will forgo mutually beneficial accumulation activities.

The hypothesis that some measure of accumulation will occur in the MSNL regime, as in the account sketched above, depends on an implicit assumption that the logic in the one-off game will not apply—that L will rationally refrain from appropriation because other C’s in future periods would not accumulate. For the indefinitely iterated game, self-restraint may be L’s best strategy. But for this to be the case—for L to exercise self-restraint indefinitely—requires assumptions about L’s perceived security of tenure and discount rate—assumptions that might be implausible under certain contingencies. And if the likelihood of these contingencies arising is high enough [given the pay-offs in Fig. 1, if the probability of expropriation is greater than 50 %], then C will rationally *not* accumulate. So it will pay L to pre-commit to the strict income tax regime, if such pre-commitment is feasible.

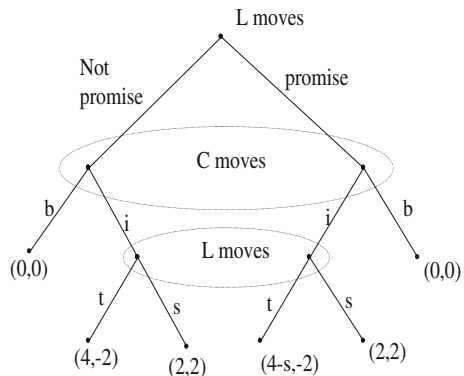
Fig. 1 The standard trust game



One thing we know *is* feasible, that is for L to make promises of type-A. For there is nothing which is available to any citizen in the state of nature that is not equally available to L and if citizens have the wherewithal to escape from the state of nature through credible promises of the type-A, then L must have access to such self-binding as well. One way to think of this issue is to postulate that promises are binding if the temptation to break them is not too great. So, the game indicated in Fig. 1 becomes the augmented game in Fig. 2, in which L has the option of making a promise; and it is common knowledge that breaking a promise costs L something—an amount s in Fig. 2. In Fig. 2, we represent the reward to L from expropriation as E , while retaining all other values as specified in Fig. 1. We can on this basis analyse the Leviathan interaction with citizens and the citizens’ emergence from the state of nature by reference to the same general analytic structure and the same two parameters. The first parameter is the value of s , which measures the force of promises made; and the second parameter is E , the value of the temptation at stake in failing to fulfil promises made. Clearly, here if $E > (s + 2)$, then promises will not be adhered to. Type-B promises are those such that this condition is satisfied. This includes specifically promises that imperil one’s life. Equally clearly, if $s > 0$, then a promise has some force and promises formally made, even without external enforcement, increase the likelihood that they will be kept across a range of possible contingencies.

Now, L may have access to a number of manoeuvres that serve to increase the likelihood that any constitutional promises made by L will be adhered to. It may be possible for L to ensure that any violation of such a promise is more public and known to be more public, so that the cost in terms of aggregate citizen response to any act of expropriation, for example, is likely to be magnified. There are, however, limits to what the Hobbesian sovereign can do by way of institutional redesign. The Hobbesian sovereign will never rationally cede ultimate power. There can be no real division of powers in the Hobbesian world—because the sovereign can always ensure his will by threat of killing the subject in question. There can be no real ceding of powers back to the people and no contingent contracts between sovereign and people. All there can be ultimately are unilateral actions on the part of the

Fig. 2 The augmented trust game



sovereign all of which must be rational, within the limits imposed by access to type-A promises. [The fact that we have modelled those type-A promises as we have allows us to recognise the moral force of the promise, the s-factor, without relinquishing the formal apparatus of rationality. The agent L chooses rationally when he keeps his promise, because the act of promising affects the pay-off structure. There may be things to be said against this formulation—but one thing to be said in its favour is that it exhibits the important property that *behaving* as morality requires is a function of the temptations involved. There is here a downward sloping demand curve for moral behaviour in the sense that the higher the opportunity cost of behaving morally the less moral conduct one will observe.]

The bottom line here is that there will be a Hobbesian Constitution, based on the limited capacity of the sovereign to bind himself through promises unilaterally made. The sovereign cannot, however, be taken to use his ‘external’ enforcement powers against himself; so the Constitution so formed has real limits from a conventional contractarian viewpoint. There can be no division or separation of powers. Power is one and lies exclusively and unalienably with the sovereign. We might therefore prefer to think of the Hobbesian Constitution as a kind of ‘covenant’, much like the old and new covenants that God is understood to make with his people in the Christian account—unilateral uncontingent outpourings of ‘grace and favour’. In the Hobbesian case, however, the grace and favour have a necessarily instrumental value to the sovereign. They enable the sovereign to enjoy a level of personal well-being that exceeds that to which he would otherwise have access.

If the Hobbesian Constitution involves no separation/division of power, what can it involve? First, for all citizens whose lifetime net contribution to the common wealth [i.e. aggregate taxable capacity] is positive, promises of protection of person and property against action from other citizens. Second, threats of swift and terrible death to any who would threaten the sovereign’s authority or person. Third, undertakings by the sovereign to exercise a measure of restraint in the exercise of his power—and specifically, to refrain from violence against citizens and their property, beyond the enforcement of the general citizen obligation to assist the sovereign in the pursuit of the sovereign’s ends and the citizen obligation to pay the tax on income imposed by the sovereign. From the sovereign’s viewpoint, all of these provisions are of the “kill not the goose that lays the golden egg” variety; and in that sense, in general keeping with the object of having the sovereign’s course go as well as possible overall. But the constitution is not externally enforceable. So its force has to lie solely on the ‘moral’ weight of promises made, and of the sovereign’s design of institutions that would make public any failure to keep promises. It seems clear to us that the idea that promises have some [but limited] weight in the state of nature is an essential feature of the Hobbesian scheme. The limits, however, are also notable. In particular, there can be no institutional arrangement that does not leave the sovereign better off; or that requires the sovereign to be subject to any measure of external enforcement (That these are not the same thing is what the trust predicament shows.).

In the Hobbesian world, the *quis custodiet ipsos custodes* challenge bites quite deeply. There can be no arrangement that violates the interests of the *custodes*; and even the advantages to be obtained from self-binding are limited to those accessible under the moral force of ordinary promises. These two constraints obviously impose quite severe limits on the scope and nature of the Hobbesian Constitution. But the constraints do not imply that the idea of a Hobbesian Constitution is a contradiction in terms.

Perhaps the spirit of the foregoing can be captured in terms of a simple diagram. On the vertical axis is the long-term annual consumption of the sovereign; and on the horizontal axis is the annual consumption of the typical citizen. The line NL indicates the notional locus of returns to both sovereign and citizen as the move out of the state of nature proceeds. It can be thought of as showing the locus of various feasible outcomes as the value of the parameter s [the moral force of type-A promises] increases from zero. On our reading of the Hobbesian logic, when s is zero, no escape from the state of nature is feasible. As s increases, two things happen. First, the promises of citizens in pledging support to the sovereign become more robust. And, second, the capacity of the sovereign to make mutually beneficial credible promises in relation to the treatment of citizens becomes more extensive (Fig. 3).

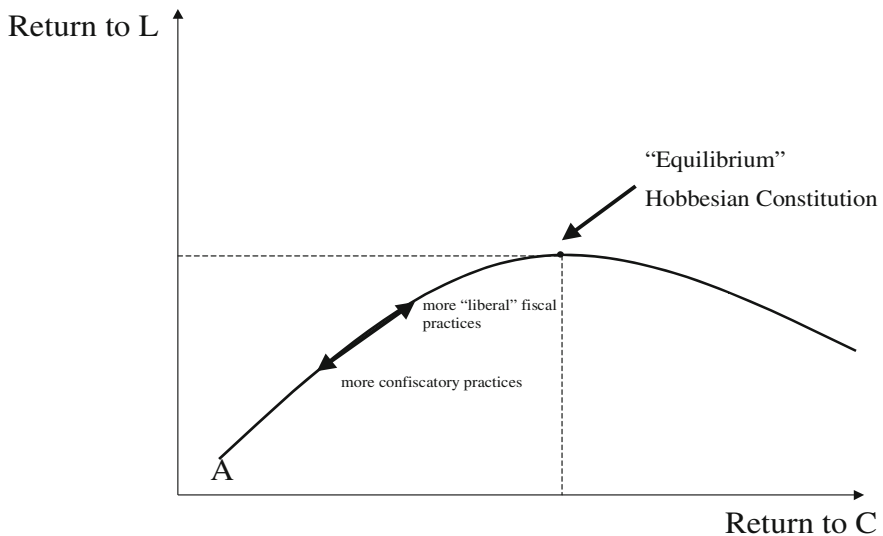


Fig. 3 The limits of sovereign and citizen common interests

5 The Hobbesian/Contractarian Contrast

The argument of this paper can be brought to a suitable conclusion by focussing on the central divergence between Hobbesian and standard contractarian approaches. The standard contractarian approach begins with a depiction of the Hobbesian state of nature as an n -person prisoners' dilemma, and focuses on the potential that all citizens would have for a much improved life under a regime of collectively enforced rules. It poses the question as to what set of rules would be ideal from the citizens' collective viewpoint and seeks to answer that question by imagining what rules would command unanimous approval, after appropriate bargaining/exchange/compromise behind an idealised 'veil of ignorance'. The Hobbesian approach, as we have depicted it here, insists on the imposition of three logical constraints in this entire exercise. First, a set of enforceable rules must be *feasible*. That is, it must be possible to establish a sovereign with the requisite powers. Hobbes takes it that this feasibility test can be met—and it is our conjecture that this is so only if there is some kind of promise/contract that can be made in the state of nature that will prove binding over some range. Second, it must be *necessary* to establish a sovereign. That is, whatever scope for agreements in the state of nature there is, such scope cannot be sufficient to allow directly binding collective contracts—since then the state of nature would not and could not be a genuine equilibrium. Third, it must be the case that the citizens have reason to believe that their lives will go better under a sovereign not subject to external enforcement than in the state of nature itself. We have sketched out an argument as to why this expectation is reasonable, based on the sovereign's capacity to gain some share of the benefits from the escape from the state of nature. Further we have argued that the same limited capacity of promise-making/keeping that enables the citizens to agree to establish an effective sovereign also enables that sovereign to commit to a limited Constitution—a 'covenant' with his people which though not subject to external enforcement nevertheless has some force, and indeed precisely the same force as required to allow the escape from the state of nature in the first place.

Suppose such a Hobbesian regime is in place. It can hardly fail to attract the attention of the contractarian that the sovereign enjoys considerable rents—rents that could in principle be redistributed to the citizenry in a variety of ways. In this respect, the contractarian is no different from the 'pretender' who lusts after the power and wealth that the sovereign possesses. But the contractarian is different in this respect—that if the gains were to be redistributed in certain ways to the citizenry, under an alternative set of institutional arrangements, there would be an increased 'pie' for redistribution. The revenue-maximising income tax regime may be better from the citizens' viewpoint than many alternatives, but it is not the best. The wealth of the nation would be yet further increased under a less-modified system of natural liberty. In this sense, however, the Hobbes-informed observer would see the contractarian as making the same mistake as the 'pretender'—the mistake of failing to recognise the relevance of the feasibility constraints—the mistake of making the good the enemy of the [feasible] best. And as Hobbes sees it,

the cost of that failure is extreme. It threatens a return to the state of nature—a state that hovers whenever citizens forget the moral force of their original promise to support the sovereign up to the limits of the cost of their own lives. The challenge that Hobbes poses for the contractarian is not to show that more liberal regimes than the Hobbesian are *desirable*, but to show how they are feasible—at least within a frame that recognises the state of nature as a relevant point of departure for constitutional analysis.

Put simply, whereas the contractarian assumes that all the gains from the exit from the state of nature are available for appropriation by the citizenry, the Hobbesian recognises that only a limited share of these gains is available. Whereas for the contractarian, the test for desirability is the conceptual agreement of all citizens as equals behind the veil of ignorance, the Hobbesian test must always include the prevailing sovereign as party and include as feasible moves only those changes that are consistent with the sovereign's interest. There are, of course, critical contractarian elements in the Hobbesian account; but those elements operate at the level of having any rules at all, rather than in settling directly the question of what the unconstrained 'ideal' set of rules might be. As we have noted, it might well be possible to offer a contractarian account of the 'reason of rules' from a point of departure other than the Hobbesian one. But once one has chosen the Hobbesian state of nature as the benchmark case of 'no rules', one seems committed to the Hobbesian logic 'all the way down'; and that commitment gives rise to an account of feasible constitutionalism that seems on its face quite different from the emerging CPE orthodoxy in certain crucial respects. Our aim in this paper has been to indicate something of what those "crucial respects" are.

References

- Brennan, G., and J.M. Buchanan. 1980. *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. New York: Cambridge University Press.
- Buchanan, J.M. 1969. *(Cost and Choice: An Inquiry in Economic Theory*. Chicago: Markham Publishing Company.
- Buchanan, J.M. (1982). The Domain of Subjective economics: Between Predictive Science and Moral Philosophy. In *Method, Process, and Austrian Economics: Essays in Honor of Ludwig von Mises*, ed. M. Kurzner, 7–20. Lexington: D.C. Heath).
- Buchanan, J.M. (1991). The Foundations for Normative Individualism. In *The Economics and the Ethics of Constitutional Order*, 221–229. Ann Harbor: University of Michigan Press.
- Buchanan, J.M. 1999 "The Relevance of Pareto Optimality"; "Positive Economics, Welfare Economics and Political Economy". In *The Collected Works*, ed. James M. Buchanan. Indianapolis: Liberty Fund.
- Buchanan, J.M., Tullock, G. 1962. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor University of Michigan Press.
- Lindahl, E. (1919). Just Taxation—A Positive Solution. In *Classics in the Theory of Public Finance*, eds. R. Musgrave and A. Peacock (1967). London: Macmillan.

Locke, J. 1689. *Two Treatises of Government*, ed. P. Laslett. Cambridge: Cambridge University Press.

Rousseau, J.J. 1762. *The Social Contract*. London: Penguin Classics.

Wicksell, K. 1896. A New Theory of Just Taxation. In *Classics in the Theory of Public Finance*, eds. R. Musgrave and A. Peacock (1967). London: Macmillan.