

In Defense of Knavish Constitutions¹

Brian Kogelmann

John Chambers College of Business and Economics, West Virginia University

Abstract: A tradition in political economy holds that constitutions should be designed under the assumption that politicians are knaves. A criticism of this position says that a constitution so designed will cause political actors to behave worse than they otherwise would. Designing a constitution for knaves *creates* knaves. I critique this argument in the current paper. I advance two claims. First, all constitutions create knaves, because the activity of politics itself creates knaves. Second, knavish constitutions better cultivate virtue when compared to constitutions that lack knavish constraints and guardrails. Put together, the two arguments imply the criticism has it exactly backwards: if you want virtuous politicians, design constitutions under the assumption that they are knaves.

Keywords: constitutional political economy; constitutional design; *homo economicus*; crowding out; moral economy; knave.

1. Introduction

A key part of the Enlightenment tradition is the idea that political constitutions should be designed under the assumption that those acting within the constitutional structure will be driven

¹ The author would like to thank Jeremy Reid, the editors of *Public Choice*, and two anonymous reviewers for helpful feedback on earlier drafts of the manuscript.

by their own self-interest rather than the public interest. Constitutions, to put it another way, should be designed under the assumption that politicians are knaves. I will call a constitution so designed a *knavish constitution*.

Knaveish constitutions would later be embraced by the *constitutional political economy* (CPE) research program. CPE is the normative branch of *public choice theory* (Kogelmann, 2022, p. 76). Public choice theory deploys the tools of economics to analyze political institutions. It thus embraces the economic theory of the person—*homo economicus*—that says persons adhere to the standard rational choice axioms; public choice theorists also typically fill in the content of persons' preferences by assuming they are driven mainly by pecuniary gain (Kogelmann, 2015, p. 456). Public choice is a social scientific enterprise whose aim is to describe and predict. CPE is a normative enterprise that asks: assuming political actors are knaves, how should we design our constitution?

Not everyone agrees constitutions should be designed under the assumption that politicians are knaves. The most prominent criticism of knaveish constitutions is that a constitution so designed will in some way cause political actors to behave worse than they otherwise would. Designing a constitution for knaves *creates* knaves. So, if you want public spirited politicians, don't craft a knaveish constitution. Call this the *crowding out argument*. The crowding out argument is a serious criticism that deserves a serious response. Defenders of knaveish constitutions have responded in the past (e.g., Brennan and Hamlin, 1995; Brennan and Hamlin, 2000, chapter 4), but there are more recent articulations of the crowding out argument that merit a fresh reply. I offer that in this paper.

Here is the structure of the argument. The next section offers an overview of knaveish constitutions and their defenders (§2). The section after explicates the crowding out argument

(§3). From here I go on the attack. I first argue that all constitutions—irrespective whether they contain knavish guardrails or not—create knaves, because the activity of politics itself creates knaves (§4). I then argue that knavish constitutions better cultivate virtue when compared to constitutions that lack knavish constraints and guardrails (§5). Put together, the two arguments imply the crowding out argument has it exactly backwards. If you want public spirited politicians, design constitutions under the assumption that they are knaves. There is a concluding section.

2. Constitutions for Knaves

The idea that constitutions should be designed for knaves is common among Enlightenment thinkers. Consider a few examples.

Political writers have established it as a maxim, that, in contriving any system of government, and fixing the several checks and controuls of the constitution, every man ought to be supposed a *knave*, and to have no other end, in all his actions, than private interest (Hume, 1987, p. 42).

... in the framing of laws, suspicion can not possibly be carried to too high a pitch ... every man ought to be presumed disposed to be guilty and endeavouring to be guilty to the purpose of legislative enactment (Bentham, 1989, p. 15).

... the very principle of constitutional government requires it to be assumed, that political power will be abused to promote the particular purposes of the holder; not because it always is so, but because such is the natural tendency of things, to guard against which is the especial use of free institutions (Mill, 2015, p. 326).

As hard as it may sound, the problem of organizing a nation is solvable even for a people comprised of devils (if only they possess understanding). The problem can be stated in this way: “So order and organize a group ... that, while their private attitudes conflict, these nonetheless so cancel one another that these beings behave publicly just as if they had no evil attitudes” (Kant, 1983, p. 124)

What kinds of misconduct are these thinkers trying to guard against with knavish constitutions? Hume was most concerned about politicians driven by the spirit of faction, what we might today call partisanship (Sagar, 2021, pp. 1098-1101). Bentham was concerned with legislators who neglect or betray their constituents’ interests (Bruno, 2017, p. 298). Opposite Bentham, Mill was concerned with politicians who were too beholden to their constituents. Right before the Mill passage cited above, he writes: “By refusing to elect any one who will not pledge himself to all their [i.e., the voters’] opinions, and even, if they please, to consult with them before voting on any important subject not foreseen, they can reduce their representative to their mere mouthpiece” (Mill, 2015, p. 326). Kant was worried about those who “are willing to use or to reject the moral law whenever it serves their own ends, and they seek others to corrupt and use as mere means” (Clohesy, 1995, p. 738).

Knavish constitutions would later be embraced by the CPE research program. While public choice theory (and economics more generally) focuses on individuals’ choices within constraints, “constitutional economics directs analytical attention to the *choice among constraints*” (Buchanan, 1999c, p. 379). To choose among constitutional constraints, we need to think through the consequences of different constitutional rules. And to do this, we need a model of human behavior. Buchanan, one of the founders of CPE, suggests that “in such a setting, there is a powerful argument that suggests the appropriateness of something akin to the *Homo*

economicus postulate for behavior” (Buchanan, 1999c, p. 392). For Buchanan, the *homo economicus* behavioral model consists of two components (Kirchgässner, 2014, section 2; Kogelmann, 2015, p. 456). First, it contains the basic rational choice postulates that say individuals have well-defined (i.e., complete, transitive, and reflexive) preferences and choose the option that best satisfies their preferences. Second, the preferences of *homo economicus* are often defined narrowly in terms of wealth maximization. To choose among constitutions, assume those acting within the constitutional structure are wealth-maximizing knaves.

Given the different ways of understanding knavish behavior, I will use the term as ecumenically as possible going forward. Many understand the relationship between politicians and their constituents in terms of a principal-agent relationship (Gailmard, 2014). The people are the principals, the politicians their agents. Politicians act knavishly when they deviate from their principal’s—the people’s—interests. There may be many reasons that motivate this deviation: because they seek pecuniary gain, because they are wrapped up in the spirit of faction, because they do not adhere to the moral law, and so on. This broad interpretation of knavish behavior is, I believe, consistent with all the thinkers just canvassed.²

Why design constitutions under the assumption that politicians are knaves? We get several arguments from those in the CPE tradition. In particular, Geoffrey Brennan and Buchanan give several arguments in defense of knavish constitutions. Importantly, they note up

² One might think it is inconsistent with Mill, who worries about politicians being turned into a “mere mouthpiece” of their constituents. I disagree. To begin, note that voters have both preferences over policy outcomes as well as beliefs about which policies best realize these outcomes (Prat, 2005). Voters, of course, may be mistaken in their beliefs (e.g., they believe immigration restrictions foster economic growth when most economists believe the opposite [Clemens, 2011]). One way of interpreting Mill’s concern is that he is worried about politicians deviating from their constituents’ outcome preferences by catering to their false beliefs about what policies best serve these preferences. Contemporary political economists call this *pandering* (Maskin and Tirole, 2004). On this interpretation of Mill’s concern, knavish politicians *are* deviating from their principals’ outcome interests by pandering to their false policy beliefs.

front that none of these arguments appeal to the accuracy of the *homo economicus* assumption. They write: “if men were genuinely as economic theory depicts them, no sort of ordered society ... would be possible ... the very existence of an ordered society casts doubt on the *Homo economicus* model of behaviour” (Brennan and Buchanan, 2000b, p. 75). So, we should design constitutions under the assumption that persons are knaves even though they are not. Why deploy an admittedly false behavioral model? Several reasons are offered:

- We should examine institutions using symmetrical behavioral assumptions. Because we already deploy *homo economicus* in our analysis of markets, we must also use it in our analysis of constitutions (Brennan and Buchanan, 2000b, pp. 77-79; Brennan and Buchanan, 2000c, pp. 56-57).
- The very point of institutional analysis is to determine the extent to which institutions can transform private vice into public benefit. Therefore, we must assume knavery when examining constitutions, to see whether constitutions can transform this knavery into public benefit (Brennan and Buchanan, 2000b, pp. 79-81; Brennan and Buchanan, 2000c, pp. 59-61).
- Because the harm imposed by bad political actors is greater than the benefit received from good political actors, we should avoid empirically accurate behavioral models and instead deploy pessimistic ones. Hence, we should assume politicians are knaves when designing constitutions (Brennan and Buchanan, 2000a; Brennan and Buchanan, 2000c, pp. 61-68).
- When agents interact with one another, the presence of some self-interested individuals forces altruistic individuals to behave in a self-interested manner to protect themselves. Since at least some politicians will be knaves, most politicians will eventually act

knavishly. To prepare for this, we must assume all persons are knaves when designing constitutions (Brennan and Buchanan, 2000c, pp. 68-73).

These are the arguments in defense of knavish constitutions among CPE practitioners. Some contemporary political theorists and philosophers also embrace knavish constitutions, offering their own defenses (e.g., Elster, 2013; Kogelmann, 2020).

As a final point before closing this section, what does a knavish constitution actually look like? There is no one answer. A constitution for knaves is one that includes constraints and guardrails to protect against opportunism. Many of these are familiar: bicameralism, separation of powers, checks and balances, enumerated and limited powers, enumerated rights, and so on. CPE practitioners have proposed several novel constitutional innovations to protect against knavery, such as a generality amendment or awarding government contracts through sortition (Kogelmann, 2022, pp. 75-79). Some have gone so far to propose that legislators should vote by secret ballot to eliminate credible commitments between politicians, lobbyists, and interest groups (Kogelmann, 2021, chapter 2). Regardless of the exact proposal, the basic idea is that we should assume power will be abused, so constitutions should include constraints and guardrails to prevent this. A *knavish constitution* is one that includes many such constraints; the more constraints there are, the more knavish it is.

3. The Crowding Out Argument

Not everyone believes constitutions should be designed under the assumption that politicians are knaves. The crowding out argument says that many politicians are not knaves, but

if you design a constitution under the assumption that they are, then they will *become* knaves. If you want public spirited politicians, don't craft a knavish constitution.

Many economists and political theorists advance the crowding out argument (e.g., Goodin, 1980, section 4; Hirschman, 1984, pp. 93-95; Frey, 1997; Pettit, 1998, pp. 74-77; Vermeule, 2003, pp. 424-425; Mansbridge, 2009, pp. 378-379; Le Grand, 2013, pp. 52-55; Vermeule, 2014, pp. 71-72; Bowles, 2016; Mansbridge, 2016, p. 57; Schmitt, 2022, section 2).

Here is a sample:

The Problem of Low Expectations is that politicians will, *ceteris paribus*, become more likely to choose their self-interest rather than exercising restraint the longer they operate within unromantic institutions (Schmitt, 2022, p. 15).

... *the expectation of knavery is self-fulfilling*: presuming officials to be knaves will tend to make officials into knaves (Vermeule, 2014, p. 71).

... if the deviant-centered strategy were implemented, and if the harsh penalties and high rewards that it would support were put in place, then this would be likely to switch many nonegocentric deliberators to egocentric mode (Pettit, 1998, p. 74).

... although institutional designers have argued in the past that institutional arrangements based on self-interest will be the most impervious to destruction, such arrangements also have the potential for crowding out public spirit (Mansbridge, 2009, p. 379).

... preventing knaves from exploiting the political system has not only advantages but also costs by crowding out civic virtue which is indispensable for an effective constitution and other public laws (Frey, 1997, p. 1052).

Though many thinkers advance the crowding out argument, they offer different accounts of *why* knavish constitutions crowd out virtue, as well as different accounts of what we should *replace* knavish constitutions with.

In what follows I shall explicate what I think are the three most plausible mechanisms in the literature explaining why knavish constitutions create knaves (§3.1). I then examine several alternatives to knavish constitutions (§3.2). In doing so I offer a detailed yet necessarily incomplete explication of the crowding out argument.

3.1 *Why do knavish constitutions create knaves?*

Why think designing a constitution under the assumption that politicians are knaves creates knaves?

Proponents of the crowding out argument often (but not always) begin by pointing to an empirical literature demonstrating that, in some cases, incentivizing persons to engage in pro-social behavior *reduces* their willingness to engage in this behavior, *contra* what most economists would predict. The most famous example is Richard Titmuss's (1997) work showing that when you pay people to donate blood, they are *less* likely to donate when compared to uncompensated donations. Another famous example involves a daycare in Haifa, Israel. To incentivize parents to pick up their children on time, the daycare started imposing fines on tardy parents. Contrary to what standard economic theory predicts, *more* parents started picking up their kids late after the fine was introduced (Gneezy and Rustichini, 2000). As a final example, most survey respondents in a small town in Switzerland proved willing to harbor nuclear waste in their town, despite the risks. Yet, when offered compensation for bearing these risks, they

were *less* willing to store the waste (Frey et al., 1996). More examples are available (e.g., Frey and Jegen, 2001; Besley and Ghatak, 2018).³

At first blush these empirical results have nothing to do with knavish constitutions and the transformative effect they supposedly have on politicians. The thought here is that whatever mechanisms explain the stultification of virtuous behavior in the above cases *also* apply to politicians when constitutions are designed under the assumption that they are knaves. That is, whatever explains why parents are less likely to pick up their kids from daycare on time when you fine them *also* explains why politicians are less likely to govern virtuously when you subject them to bicameralism, checks and balances, and other safeguards. Hence, the thesis that knavish constitutions create knaves is *inspired* by empirical work, but ultimately proceeds on theoretical grounds. Let's take a look at some of these mechanisms.

One explanatory mechanism says introducing incentives provides information, and this information can have a negative impact on virtuous motivation. More specifically, when a principal imposes an incentive on an agent, the agent learns something about the principal's beliefs, which can affect her motivation. The agent might learn, for instance, that the principal views her as untrustworthy, so she responds by acting in an untrustworthy manner. As Bruno Frey describes it:

An intervention from the outside undermines the actor's intrinsic motivation if it carries the notion that the actor's intrinsic motivation is not acknowledged. The person affected feels that his or her competence is not appreciated which leads to an impaired self-esteem, resulting in a reduced intrinsic motivation (Frey, 1997, p. 1045).⁴

³ For criticism of this empirical work, see Esteves-Sorenson and Broce (2022).

⁴ See also Bowles (2016, pp. 86-89).

Consider an example. In the early 2000's, the Boston fire commissioner noticed a suspicious number of sick call-ins on Mondays and Fridays, so he ended the department's policy of unlimited sick days, imposing a limit of 15 per year. Sick days in excess of 15 would result in docked pay. Contrary to what one might expect, the number of firemen calling in sick on holidays increased *tenfold* from the previous year (Bowles, 2016, p. 9). Why? The most plausible answer is that by introducing an incentive, the commissioner informed his employees of what he thought about them. By limiting sick days, the commissioner "conveyed the information that he did not trust that the firemen were doing their very best to come to work" (Bowles, 2016, p. 87). Having learned their boss did not trust them, firemen's motivation to show up and work at inconvenient times decreased.

How does this mechanism apply to knavish constitutions? The idea is that by designing constitutions with knavish constraints—bicameralism, checks and balances, separation of powers, and the like—politicians are informed that their principals (the people) do not trust them to govern virtuously. Rebuffed by this information, politicians' motivation to govern virtuously attenuates, and they become knaves.

A second explanatory mechanism appeals to what is sometimes called *moral disengagement*. The basic idea is that our moral motivation is upregulated in some environments, downregulated in others. The presence of incentives that appeal to our self-interest is one factor that can downregulate our moral motivation, resulting in selfish behavior. Samuel Bowles writes:

Incentives may cause crowding out for another reason, one less familiar to economists. In most situations, people look for cues for appropriate behavior, and incentives provide these cues. A plausible explanation of some of the framing effects of incentives is that it occurs because market-like incentives trigger what psychologists term "moral

disengagement,” a process that occurs because “people can switch their ethicality on and off” (Bowles, 2016, pp. 89-90).⁵

For example, when we engage in market exchange our moral motivation is downregulated because the monetary incentives that we confront appeal to our self-interest. Other environments upregulate our moral motivation, such as when we engage in personal relationships, when we witness suffering, when we see others acting charitably, and so on. This mechanism, in my view, best explains the blood donation case. Something about uncompensated blood donations upregulates persons’ moral motivation, inducing them to act altruistically; introducing monetary compensation downregulates their moral motivation, leading them to act selfishly.

Applied to knavish constitutions, the idea is that knavish constraints downregulate politicians’ moral motivation, because these constraints appeal to their opportunistic impulses. Just as paying people to donate blood downregulates their moral motivation (because doing so appeals to their self-interest), so does introducing bicameralism, separation of powers, and the like for politicians.

A third mechanism begins with the premise that we don’t like being controlled by others. When control is exercised over us, we sometimes respond with intransigence, doing the exact opposite of what the controller desires. Writes Frey:

... when individuals perceive the external intervention to be ‘controlling’ in the sense of reducing the extent to which they can determine actions by themselves, intrinsic motivation is substituted by extrinsic control ... They rationally reduce the motivational factor under their control, that is intrinsic motivation (Frey, 1997, p. 1045).⁶

⁵ See also Pettit (1998, p. 75).

⁶ See also Le Grand (2013, p. 53); Bowles (2016, pp. 97-103).

To put it another way, when people try to control our conduct by imposing incentives on us, we sometimes respond by doing the exact opposite of what they want us to do as a form of resistance.

Applying this mechanism to constitutions: politicians, incensed by the confinements imposed on them by a knavish constitution, respond by doing the exact opposite of what the people (the designers of the knavish constitution) want. The people impose knavish constraints to induce virtuous behavior. But because they are upset that their autonomy has been restricted, politicians do the exact opposite and act knavishly.

3.2 What should we replace knavish constitutions with?

What do proponents of the crowding out argument suggest we replace knavish constitutions with? After all, nothing in the crowding out argument denies principal-agent problems are endemic to politics. The crowding out argument only says resolving these dilemmas via knavish constraints is likely to yield even worse opportunism. Principal-agent problems must be resolved in some other way, though.

Some proponents of the crowding out argument suggest we focus less on constraining knavish politicians and more on selecting politicians that don't need so much constraining in the first place. Jane Mansbridge, for instance, says voters should select a representative who "already has self-motivated, exogenous reasons for doing what the principal wants" (Mansbridge, 2009, p. 269). In similar spirit, after offering his normative theory of the principled politician, Sam Schmitt recommends that his theory "serve as an additional criterion for voters" (Schmitt, 2022, p. 20).

Frey thinks we need to trust politicians more. He writes: “a constitution must thus reflect trust towards citizens and politicians in order to safeguard their civic virtue, and at the same time put limits on politicians in order to prevent them from exploiting the ordinary citizens” (Frey, 1997, p. 1049). I interpret Frey as saying that politicians should be afforded more discretion and be subject to fewer knavish constraints than the thinkers canvassed in §2 would recommend; they still, however, should be subject to some constraints.⁷ I interpret Adrian Vermeule as saying something similar. Vermeule comes out against the view that says “constitutional rules should above all entrench precautions against the risks that official action will result in dictatorship or tyranny, corruption and official self-dealing, violations of the rights of minorities, or other political harms” (Vermeule, 2014, p. 11). Instead, Vermeule adopts what he calls *optimizing constitutionalism* which recognizes both the harms of knavish constraints (such as motivational crowding) as well as the benefits of vigorous official action (Vermeule, 2014, pp. 10-14). Though Vermeule is reluctant to make specific institutional proposals (Vermeule, 2014, pp. 84-85), I understand his ideal constitutional structure as one that affords more discretion to politicians when compared to a knavish constitution.

Continuing, Philip Pettit says we prevent opportunism and avoid crowding out by not only screening for good politicians as Mansbridge and Schmitt suggest (Pettit, 1998, pp. 79-81), but also by sanctioning and constraining politicians in gentler ways that prime them to be ethical. Force public officials, he recommends, to defend their votes publicly. If they cannot muster a compelling defense, “they will lose face with the others; they will look silly or prejudiced”

⁷ More specifically, instead of bicameralism, separations of powers, checks and balances, and other commonly embraced knavish constraints, Frey recommends giving citizens the power to reverse legislative decisions via referenda and force the passage of laws opposed by their representatives via initiatives. A constitution with these constraints on the legislature and no others is one that affords significant legislative discretion.

(Pettit, 1998, p. 84).⁸ According to Bowles, “good policies and constitutions are those that support socially valued ends not only by harnessing self-interest but also by evoking, cultivating, and empowering public-spirited motives” (Bowles, 2016, p. 222). How do we do this? Similar to Pettit, Bowles recommends minor constraints and sanctions that come packaged with a “moral lesson” (Bowles, 2016, pp. 200-205). He also suggests that instead of constraining bad behavior we reward good behavior (Bowles, 2016, pp. 188-189, p. 219).

In contrast to knavish constitutions, call an *anti-knavish constitution* one that (i) contains comparatively few knavish constraints, and (ii) attempts to resolve the principal-agent problems endemic to politics with some combination of the mechanisms described in this subsection. The crowding out argument says anti-knavish constitutional structures will be occupied by more virtuous politicians when compared to knavish ones. This is because knavish constitutions create knaves. Anti-knavish constitutions prevent opportunism without creating knaves.

4. Politics Crowds Out Virtue

The crowding out argument says knavish constitutions create knaves. As such, we are better off with an anti-knavish constitution which does not create knaves; such a constitution is our best hope for preserving whatever public spirit exists among politicians. I don’t buy the argument and launch a two-pronged attack in this section and the next. In this section I argue that anti-knavish constitutions create knaves. In the next section I argue that knavish constitutions are more likely to cultivate virtue than anti-knavish ones. The implication is that the crowding out

⁸ Pettit recognizes that these gentle, ethics-priming constraints might not always work. If they do not, then more significant sanctions and constraints can be applied, but only in an escalating fashion, after the gentler methods fail (Pettit, 1998, pp. 85-87). Pettit does not give a clear example of how this would work.

argument has it exactly backwards. If you want public spirited politicians, design your constitution under the assumption that politicians are knaves.

The central claim of this section is that anti-knavish constitutions create knaves. This has nothing to do with their features and everything to do with politics as an activity. Regardless of the constitutional structure it occurs within, *politics creates knaves*. To defend this claim, I shall argue that the three crowding out mechanisms highlighted in §3.1 are present in ordinary democratic politics irrespective of constitutional structure. Whether we have a knavish or anti-knavish constitution, virtue will be crowded out, because politics crowds out virtue.

Recall the first mechanism: when principals incentivize agents, agents learn their principals do not trust them, which affects their motivation. Applied to knavish constitutions, politicians are informed that their principals (the people) do not trust them to govern virtuously when knavish constraints are imposed on them. So, they respond by acting knavishly.

Even without knavish constraints, however, politicians will know that many do not trust them. There are still elections in anti-knavish constitutional structures, so successful candidates will have to deal with having their character and integrity questioned in debates, town halls, and negative attack ads. There is still a free press in anti-knavish constitutional structures, so politicians will constantly be monitored and hounded by journalists. Politicians must still travel to their home districts and deal with angry constituents. Public opinion polling still exists, and politicians will be able to access this information, learning just how little the public trusts them. If learning one is not trusted is sufficient to induce knavery, then we should expect politicians in both knavish and anti-knavish regimes to become knaves, for the democratic process often reveals to politicians just how little faith the people have in them.

Turn now to the second mechanism: when an agent's self-interest is appealed to (as it is, for instance, in markets) their moral motivation is downregulated, dissipating altruistic behavior. Applied to knavish constitutions, politicians' moral motivation is downregulated because knavish constraints are designed with their baser motives in mind. Their moral motivation downregulated, politicians become knaves.

Regardless of constitutional structure, though, politicians will frequently have their self-interest appealed to. Indeed, Bowles argues that market exchange downregulates our moral motivation, but politics—according to the public choice paradigm—is just another medium of exchange (Buchanan, 1999b, p. 50). Even in anti-knavish constitutional structures, politicians will be approached by lobbyists and interest groups who offer money and support in exchange for legislative favors (Morton and Cameron, 1992, p. 88). Bills will only get passed when legislators logroll or exchange votes with one another (Buchanan and Tullock, 1999, pp. 135-136). We can even understand the relationship between politicians and their constituents as an exchange relationship: in exchange for votes, politicians supply policies the people want, even when they know these policies are misguided (Maskin and Tirole, 2004). If ordinary market exchange downregulates moral motivation, then surely political exchange will as well. Political exchange is part of the democratic process, however, and is thus present in both anti-knavish and knavish constitutional structures.

Consider now the third mechanism: when people perceive restrictions placed on their autonomy, they sometimes respond by doing the exact opposite of what is wanted of them. In the case of knavish constitutions, the people impose knavish constraints on politicians to induce virtuous behavior. Incensed by this, politicians do the exact opposite of what the people want and become knaves.

Politicians in anti-knavish constitutional structures will see their autonomy restricted. They may not be blocked by bicameralism or judicial review or enumerated and limited powers, but there are other roadblocks that prevent politicians in these regimes from doing what they want. Most obviously, facing regular elections restricts what politicians can do; some policies might seem like a good idea, but if their constituents don't support these policies, then politicians cannot pursue them. In response, it might be argued that serving the people's interests does not restrict politicians' autonomy, because politicians want to serve their constituents. Even if this is true, there are several other groups who restrict politicians' autonomy. One is parties. Parties have goals that conflict with members' goals; parties have many ways of disciplining members to ensure they do as the party desires (Kam, 2009; Pearson, 2015). Moreover, politicians are often beholden to special interest groups. Because they have a comparative advantage in monitoring and pressuring politicians (Lohman, 1998), special interest groups exercise significant control over how politicians vote. All these factors suggest that politicians will see their autonomy restricted in anti-knavish as well as knavish constitutional structures.

To sum up the argument: the mechanisms that purportedly crowd out virtuous behavior in knavish constitutional structures are also present in anti-knavish regimes, because these mechanisms are present wherever ordinary democratic politics is present. Thus, both institutional environments crowd out virtue. Anti-knavish constitutions, just like knavish ones, create knaves.

An objection to my argument says that even though these three mechanisms are present in anti-knavish constitutional structures, they are present to an even greater degree in knavish ones. Politicians receive information that the people do not trust them in anti-knavish regimes, but they receive even more of this information in knavish constitutions. Moral disengagement happens in anti-knavish constitutional structures but is carried to an even higher pitch in knavish

ones. Anti-knavish constitutions restrict politicians' autonomy but not to as great a degree as knavish constitutions do. Anti-knavish constitutions might produce knaves, so the objection goes, but knavish constitutions create even worse knaves, for in these constitutional structures the three mechanisms are more pronounced.

I do not think this objection succeeds. I have three responses to it.

First, it's not obvious knavish constraints always embody our three mechanisms.

Consider the first mechanism: learning your principal distrusts you may lead you (their agent) to act in an untrustworthy manner. Do knavish institutions like bicameralism and judicial review communicate to politicians the people distrust them? That depends on *when* the constitution was designed and *who* was involved. A legislator in the United States cannot infer the people distrust her from the fact that the executive can veto her branch's legislation, or the judiciary declare it unconstitutional. She can only infer that James Madison and company distrusted people like her when they penned the Constitution in 1787.

As another example, it's not obvious institutions such as limited and enumerated powers or enumerated rights lead to moral disengagement. Bowles is clear that moral disengagement occurs when persons are involved in exchange relations, but enumerated powers and rights do not obviously introduce exchange. So, as a first rejoinder to the objection: we should be skeptical knavish constraints induce further knavery above and beyond what is induced by ordinary democratic politics because it's unclear the extent to which knavish constraints are actual embodiments of the relevant mechanisms.

Second, the relationship between knavery and our three mechanisms is likely at least somewhat inelastic. At some threshold, learning that others distrust you likely influences your

behavior, but from this it does not follow that *every* report of distrust influences your behavior incrementally. To put it another way, learning that 1,000 people distrust you will probably lead you to act in an untrustworthy manner, but learning about a 1,001st distruster does not mean you will sink further into treachery.

The same applies, *mutatis mutandis*, to the other mechanisms. Being locked up in prison with one hour of recreation time a day is a serious restriction on one's autonomy that likely induces behavioral changes. I am skeptical, however, that reducing the one hour of recreation to 55 minutes will induce an even worse behavioral response. As such, it does not follow that introducing knavish constraints will result in worse behavior above and beyond what we see in anti-knavish constitutions. The presence of the three mechanisms in ordinary democratic politics may do all the damage there is to be done.

Third, knavish constraints may *counteract* our mechanisms in some cases. Consider an example. A politician is beholden to a special interest group who wants her to pass a bill restricting a minority group's religious freedom. The politician does not want to do this. Suppose the politician operates in a knavish constitutional structure which includes an enumerated right to the free exercise of religion. The politician can shrug her shoulders at the interest group: she would help them if she could, but the constitution prevents her. The interest group backs off as a result. In an anti-knavish constitutional structure, the politician may lack such recourse if there is no enumerated right to free exercise of religion. Here, the interest group's pressure will be suffocating.

In this case, it seems the politician has more autonomy in the knavish regime when compared to the anti-knavish one *precisely because* of the knavish constraints. Just as being tied to the mast grants Odysseus the freedom to listen to the sirens, knavish constraints can likewise

increase politicians' autonomy to do as they wish, at least in some cases. So, knavish constraints can sometimes interact with our mechanisms in complex ways that *reduce* the mechanisms' influence. We thus cannot conclude that the presence of knavish constraints induces extra knavery above and beyond what we find in ordinary democratic politics.

Let me close and summarize this section. Virtue will be crowded out in both knavish and anti-knavish constitutional structures because the activity of politics crowds out virtue. Moreover, there are reasons to doubt knavish constitutions will crowd out virtue to a greater degree than anti-knavish ones.

5. Knavish Constitutions Cultivate Virtue

My argument in the prior section seems to suggest that we should be indifferent between knavish and anti-knavish constitutions when it comes to preserving virtuous politicians. This is incorrect. While both institutional environments crowd out virtue, knavish constitutions partially counteract this tendency by *cultivating* virtue. There are forces in knavish constitutional structures that encourage politicians to act knavishly, but there are also forces in these regimes that encourage the development of virtue. We are less likely to find such forces in anti-knavish constitutions. Hence, knavish constitutions will be populated by more virtuous politicians than anti-knavish ones. This gives us reason to favor knavish constitutions.

My thesis is in direct conflict with Sam Schmitt's, so I want to briefly look at his work before I jump into my argument. Schmitt proposes a mechanism different than the three already examined to explain why knavish constitutions create knaves. The mechanism Schmitt develops is grounded in Buchanan's "Natural and Artifactual Man." According to Buchanan, "We are, and

will be, at least in part, that which we make ourselves to be. We construct our own beings ... We are artifactual, as much like the pottery sherds that archaeologists dig up as like the animals whose fossils they also find” (Buchanan, 1999a, p. 247). Humans, in other words, can decide what kinds of persons they want to become by choosing to develop certain skills, character traits, and preferences.

Such choices do not happen in a vacuum, however. Our environment—and, in particular, the institutions we live under—influences who we decide to become. As Schmitt phrases it: “individuals within institutions are shaped in part by their own will, but are always being shaped by their institutional context as well” (Schmitt, 2022, p. 10). Most obviously, our environment influences the skills we choose to develop. Someone who grows up in an agrarian economy develops different skills than someone who grows up in a knowledge economy. But our environment also influences other traits and characteristics we might choose to develop, such as our morals. Those who reside in corrupt, violent places might have reason to become ruthless; those lucky enough to live in peaceful, well-governed countries will likely develop different moral codes.

Because our institutional environment influences who we decide to become, knavish constitutions influence the character traits politicians choose to develop. More specifically:

... where personal restraint is not often needed, agents will have *pro tanto* reason to invest in some other capacity. This means politicians will often lack the ability to take actions which they believe to be right but which do not comport with their self-interest (Schmitt, 2022, p. 10).

The idea here is that knavish constitutions discourage politicians from developing prudence, because that trait is not needed in such regimes. Why learn self-restraint when you are already restrained? To put it another way, “a decreased need for personal restraint can lead to individual weakness of personal restraint” (Schmitt, 2022, p. 10). Hence, politicians in knavish constitutional structures will not develop virtuous character traits.

I disagree. When it comes to the development of virtue, Schmitt views knavish constraints as akin to crutches: if you use them too much, you end up walking with a limp. I see knavish constraints as more like training wheels, in that using them helps develop the skills needed to ride a bike. This is because virtue is developed through *habituation*, and knavish constraints, I shall argue, habituate virtue. Because virtue is difficult to measure, it would be hard to demonstrate this thesis empirically. So, in what follows I present a theoretical argument for why knavish constitutions cultivate virtue. My argument is consistent, however, with a literature establishing that morality is at least partly endogenous to institutions (e.g., Alesina and Glaeser, 2004, chapter 7; Alesina and Fuchs-Schündeln, 2007; Tabellini, 2008). My argument can be interpreted as a hypothesis concerning how certain kinds of constitutional structures (i.e., knavish ones) influence politicians’ ethical behavior.

On my side is Aristotle. In the *Nicomachean Ethics*, Aristotle writes that “moral virtue comes about as a result of habit” (Aristotle, 2001, p. 952). He continues:

... the virtues we get by first exercising them, as also happens in the case of the arts as well. For the things we have to learn before we can do them, we learn by doing them, e.g., men become builders by building and lyre-players by playing the lyre; so too we become just by doing just acts, temperate by doing temperate acts, brave by doing brave acts (Aristotle, 2001, p. 952).

The key claim here is that we become virtuous by doing virtuous acts, a claim that Aristotle is not alone in defending (e.g., Annas, 2013, chapter 3). For example, you don't give to charity because you are magnanimous; you develop magnanimity by repeatedly giving to charity. You don't risk your life in battle because you are courageous; rather, you develop courage by repeatedly risking your life in battle.

Assume Aristotle and his followers are correct that virtue is acquired through habituation; what does this have to do with knavish constitutions? Note first that we often habituate virtue by subjecting people to rules (Chappell, 2014, pp. 83-85; Reid, 2019, pp. 89-90). For example, if you want to habituate the virtue of abstemiousness in your children, you might require they finish their vegetables at every meal and only allow them dessert once a week. By following these rules, the hope is that they develop a good set of habits and grow up to exercise moderation in their food and drink choices. As another example, if you want to inculcate work ethic and discipline in your children, you might require they finish all their homework before watching TV or playing video games; you might also require they perform chores around the house and get a job once they reach a certain age. By following these rules, the hope is that they acquire a good set of habits and grow up to be hardworking and disciplined individuals.

So, acquiring virtue requires habituation and habituation happens through rule-following. From here it is a quick jump to the conclusion that knavish constitutions inculcate virtue, for *knaveish constraints subject politicians to rules that, if consistently followed, inculcate virtuous habits in them*. Aristotle recognized this. He states that good laws inculcate virtuous habits. As such, what makes a constitution good is its inculcation of the right habits through imposition of the right rules. In his words:

... we become just by doing just acts, temperate by doing temperate acts, brave by doing brave acts.

This is confirmed by what happens in states; for legislators make the citizens good by forming habits in them, and this is the wish of every legislator and those who do not affect it miss their mark, and it is in this that a good constitution differs from a bad one (Aristotle, 2001, p. 952).

The idea is that just as children grow up to be virtuous if they are required to eat their vegetables and do their homework, people become virtuous if they are required to follow the kinds of laws that inculcate virtuous habits. Applied to politicians, they develop virtuous habits if the right kinds of rules are imposed on them. Knavish constraints, I contend, are the right kinds of rules.

Let's consider a few examples. One virtue we want politicians to possess is a willingness to compromise (Wendt, 2016). How is this virtue acquired? If Aristotle and his followers are right, then it is acquired by making compromise a habit, and this is done by imposing rules that force compromise. Knavish constitutions do this. One knavish constraint is bicameralism. Bicameralism forces legislative chambers to bargain and hence compromise with one another (Tsebelis and Money, 1997, p. 74, p. 85). Forced to negotiate with the Senate, members of the House of Representatives make compromise a habit, and thereby develop the relevant virtue. Another knavish constraint is an executive veto. Like bicameralism, this guardrail forces the legislative and executive branches to bargain and hence compromise with each other (Cameron and McCarty, 2004). By being forced to negotiate and make concessions, knavish constraints make compromise a habit.

Another virtue we want politicians to possess is legislative restraint. It is the job of legislators to pass laws, but a good legislator also knows when not to pass a law. If Aristotle and his followers are correct, then this virtue is attained by making legislative restraint a habit, and this is done by applying rules that force legislators to legislate in a restrained manner. Knavish constitutions supply these rules. Consider limited and enumerated powers as well as enumerated rights. By restricting the kinds of issues politicians can legislate on, these guardrails make legislative restraint a habit and thus inculcate the virtue of restraint.

As a final example, we want politicians who respect and revere the constitution. This virtue is acquired, according to the Aristotelian theory of virtue, by making respect for the constitution a habit, which happens by subjecting politicians to rules that force them to respect and revere the constitution. Knavish constitutions do this through judicial review. To avoid having their laws struck down, politicians are forced to think carefully about the constitutionality of the laws they pass. This forced reflection makes respect and reverence for the constitution a habit, inculcating the relevant virtue.

Other examples are available of how knavish constitutions habituate virtue through rule-following, but I don't want to belabor the point. To be clear, I am not arguing that politicians in knavish regimes will be angels. Humans are still humans, and the activity of politics (I argued in the last section) crowds out virtue. I advance the more modest claim that in knavish constitutional structures there are some forces that counteract the tendency for politics to devolve into complete knavery.

Anti-knavish constitutions, on the other hand, do not inculcate virtue as well as knavish constitutions do. Recall the definition of an anti-knavish constitution; it is one that *(i)* contains comparatively few knavish constraints, and *(ii)* attempts to resolve the principal-agent problems

endemic to politics with alternative mechanisms, such as political selection, increased trust and discretion, ethical priming, and so on. Because anti-knavish constitutions do not subject politicians to as many constraints as knavish constitutions do, virtue is less likely to be habituated. Just as subjecting children to too few rules will probably result in vice, so will subjecting politicians to too few knavish guardrails.

I can now summarize this section's argument. Knavish and anti-knavish constitutions both crowd out virtue, but only one institutional environment works to counteract this tendency by cultivating virtue. Knavish constitutions cultivate virtue. Virtue is inculcated through habituation, and this is accomplished by subjecting persons to the right kinds of rules. By subjecting politicians to many constraints and guardrails, knavish constitutions require politicians to follow rules that make virtue a habit. Because they impose fewer constraints and guardrails, anti-knavish constitutions are less likely to make virtue a habit.

6. Conclusion

We want virtuous politicians. An often-repeated line of argument says that if we design constitutions under the assumption that politicians are knaves—a common practice among Enlightenment thinkers and practitioners of constitutional political economy—then politicians will *become* knaves. I criticized this popular argument. All constitutions crowd out virtue because the activity of politics crowds out virtue, but knavish constitutions counteract some of this crowding out by inculcating virtue. If we want virtuous politicians, then we should follow Hume's advice that "in contriving any system of government ... every man ought to be supposed a *knave*."

Works Cited

- Annas, Julia. (2013). *Intelligent Virtue*. Oxford University Press.
- Alesina, Alberto and Edward Glaeser. (2004). *Fighting Poverty in the US and Europe: A World of Difference*. Oxford University Press.
- Alesina, Alberto and Nicola Fuchs-Schündeln. (2007). “Good-Bye Lenin (or Not)?: The Effect of Communism on People’s Preferences.” *American Economic Review* 97: 1507-1528.
- Aristotle. (2001). *The Basic Works of Aristotle*, edited by Richard McKeon and introduction by C.D.C. Reeve. The Modern Library.
- Bentham, Jeremy. (1989). *The Collected Works of Jeremy Bentham: First Principles of Preparatory to Constitutional Code*, edited by Philip Schofield. Oxford University Press.
- Besley, Timothy and Maitreesh Ghatak. (2018). “Prosocial Motivation and Incentives.” *Annual Review of Economics* 10: 411-438.
- Bowles, Samuel. (2016). *The Moral Economy: Why Good Incentives Are No Substitute for Good Citizens*. Yale University Press.
- Brennan, Geoffrey and James M. Buchanan. (2000a). “Predictive Power and the Choice among Regimes.” In *The Collected Works of James M. Buchanan, Vol. 12: Economic Inquiry and Its Logic*: 87-109. Liberty Fund.
- Brennan, Geoffrey and James M. Buchanan. (2000b). “The Normative Purpose of Economic ‘Science’: Rediscovery of an Eighteenth Century Method.” In *The Collected Works of James M. Buchanan, Vol. 12: Economic Inquiry and Its Logic*: 70-86. Liberty Fund.

Brennan, Geoffrey and James M. Buchanan. (2000c). *The Reason of Rules: Constitutional Political Economy*. Liberty Fund.

Brennan, Geoffrey and Alan Hamlin. (1995). "Economizing on Virtue." *Constitutional Political Economy* 6: 35-56.

Brennan, Geoffrey and Alan Hamlin. (2000). *Democratic Devices and Desires*. Cambridge University Press.

Bruno, Jonathan R. (2017). "Vigilance and Confidence: Jeremy Bentham, Publicity, and the Dialectic of Political Trust and Distrust." *American Political Science Review* 111: 295-307.

Buchanan, James M. (1999a). "Natural and Artifactual Man." In *The Collected Works of James M. Buchanan, Vol. 1: The Logical Foundations of Constitutional Democracy*: 246-259. Liberty Fund.

Buchanan, James M. (1999b). "Politics without Romance: A Sketch of Positive Public Choice Theory and Its Normative Implications." In *The Collected Works of James M. Buchanan, Vol. 1: The Logical Foundations of Constitutional Democracy*: 45-59. Liberty Fund.

Buchanan, James M. (1999c). "The Domain of Constitutional Economics." In *The Collected Works of James M. Buchanan, Vol. 1: The Logical Foundations of Constitutional Democracy*: 377-395. Liberty Fund.

Buchanan, James M. and Gordon Tullock. (1999). *The Calculus of Consent: The Logical Foundations of Constitutional Democracy*. Liberty Fund.

Cameron, Charles and Nolan McCarty. (2004). "Models of Vetoes and Veto Bargaining." *Annul Review of Political Science* 7: 409-435.

Chappell, Timothy. (2014). "Virtue and Rules." In *The Handbook of Virtue Ethics*, edited by Stan van Hoof: 76-6. Taylor and Francis.

Clemens, Michael A. (2011). "Economics and Emigration: Trillion-Dollar Bills on the Sidewalk?" *Journal of Economic Perspectives* 25: 83-106.

Clohesy, William. (1995). "A Constitution for a Race of Devils." *Proceedings of the Eight International Kant Congress 2*: 733-741.

Elster, Jon. (2013). *Securities Against Misrule: Juries, Assemblies, Elections*. Cambridge University Press.

Esteves-Sorenson, Constança and Robert Broce. (2022). "Do Monetary Incentives Undermine Performance on Intrinsically Enjoyable Tasks? A Field Test." *Review of Economics and Statistics* 104: 67-84.

Frey, Bruno S. (1997). "A Constitution for Knaves Crowds out Civic Virtues." *The Economic Journal* 107: 1043-1053.

Frey, Bruno S. and Reto Jegen. (2001). "Motivation Crowding Theory." *Journal of Economic Surveys* 15: 589-611.

Frey, Bruno S., Felix Oberholzer-Gee, and Reiner Eichenberger. (1996). "The Old Lady Visits Your Backyard: A Tale of Morals and Markets." *Journal of Political Economy* 104: 1297-1313.

Gailmard, Sean. (2014). "Accountability and Principal-Agent Theory." In *The Oxford Handbook of Public Accountability*, edited by Mark Bovens, Robert E. Goodin, and Thomas Schillemans: 90-105. Oxford University Press.

Gneezy, Uri and Aldo Rustichini. (2000). "A Fine is a Price." *Journal of Legal Studies* 29: 1-17.

- Goodin, Robert E. (1980). "Making Moral Incentives Pay." *Policy Sciences* 2: 131-145.
- Hirschman, Albert O. (1984). "Against Parsimony: Three Ways of Complicating Some Categories of Economic Discourse." *American Economic Review* 74: 89-96.
- Hume, David. (1987). *Essays: Moral, Political, and Literary*, edited by Eugene F. Miller. Liberty Fund.
- Kam, Christopher J. (2009). *Party Discipline in Parliamentary Politics*. Cambridge University Press.
- Kant, Immanuel. (1983). *Perpetual Peace and Other Essays*. Hackett Publishing.
- Kirchgässner, Gebhard. (2014). "The Role of Homo Oeconomicus in the Political Economy of James Buchanan." *Constitutional Political Economy* 25: 2-17.
- Kogelmann, Brian. (2015). "Modeling the Individual for Constitutional Choice." *Constitutional Political Economy* 26: 455-474.
- Kogelmann, Brian. (2020). "The Future of Political Philosophy: Non-Ideal and West of Babel." *Review of Austrian Economics* 33: 237-252.
- Kogelmann, Brian. (2021). *Secret Government: The Pathologies of Publicity*. Cambridge University Press.
- Kogelmann, Brian. (2022). "Public Choice and Political Equality." In *Wealth and Power: Philosophical Perspectives*, edited by Michael Bennett, Huub Brouwer, and Rutger Claassen: 67-84. Routledge.
- Le Grand, Julian. (2013). *Motivation, Agency, and Public Policy: Of Knights and Knaves, Pawns and Queens*. Oxford University Press.

- Lohman, Susanne. (1998). "An Information Rationale for the Power of Special Interests." *American Political Science Review* 92: 809-827.
- Mansbridge, Jane. (2009). "A 'Selection Model' of Political Representation." *Journal of Political Philosophy* 17: 369-398.
- Mansbridge, Jane. (2016). "A Contingency Theory of Accountability." In *The Oxford Handbook of Public Accountability*, edited by Mark Bovens, Robert E. Goodin, and Thomas Schillemans: 55-68. Oxford University Press.
- Maskin, Eric and Jean Tirole. (2004). "The Politician and the Judge: Accountability in Government." *American Economic Review* 94: 1034-1054.
- Morton, Rebecca, and Charles Cameron. (1992). "Elections and the Theory of Campaign Contributions." *Economics and Politics* 4: 79-108.
- Mill, John Stuart. (2015). *On Liberty, Utilitarianism, and Other Essays*, edited by Mark Philp and Frederick Rosen. Oxford University Press.
- Pearson, Kathryn. (2009). *Party Discipline in the U.S. House of Representatives*. University of Michigan Press.
- Pettit, Philip. (1998). "Institutional Design and Rational Choice." In *The Theory of Institutional Design*, edited by Robert E. Goodin: 54-89. Cambridge University Press.
- Prat, Andrea. (2005). "The Wrong Kind of Transparency." *American Economic Review* 95: 862-877.
- Reid, Jeremy. (2019). "Virtue, Rule-Following, and Absolute Prohibitions." *Journal of the American Philosophical Association* 5: 78-97.

Sagar, Paul. (2021). "Between Virtue and Knavery: Hume and the Politics of Moderation." *Journal of Politics* 83: 1097-1113.

Schmitt, Sam. (2022). "The Problem of Low Expectations and the Principled Politician." *Economics & Philosophy* (First View): 1-22.

Tabellini, Guido. (2008). "Institutions and Culture." *Journal of the European Economic Association* 6: 255-294.

Titmuss, Richard M. (1997). *The Gift Relationship: From Human Blood to Social Policy*, edited by Ann Oakley and John Ashton. The New Press.

Tsebelis, George and Jeannette Money. (1997). *Bicameralism*. Cambridge University Press.

Vermeule, Adrian. (2003). "Hume's Second-Best Constitutionalism." *University of Chicago Law Review* 70: 421-437.

Vermeule, Adrian. (2014). *The Constitution of Risk*. Cambridge University Press.

Wendt, Fabian. (2016). *Compromise, Peace and Public Justification: Political Morality Beyond Justice*. Palgrave Macmillan.