Quarterly Journal of Political Science, 2014, 9: 45–86

Designing Checks and Balances*

Tiberiu Dragu[1], Xiaochen Fan[2], and James Kuklinski[3]

[1] *New York University, New York, NY, USA; tiberiu.dragu@nyu.edu*
[2] *New York University, New York, NY, USA; xiaochen.fan@nyu.edu*
[3] *University of Illinois at Urbana-Champaign, Champaign, IL, USA; kuklinsk@illinois.edu*

ABSTRACT

Essential to the Madisonian conception of checks and balances is the division of policy-making authority among multiple actors such that each can veto proposed changes in policy. We use a mechanism design approach to analyze checks and balances institutions. We show that checks and balances institutions in which the most preferred policy of the more moderate player is the unique equilibrium outcome are the only checks and balances institutions that are strategy-proof, efficient, and responsive. Our analysis facilitates a comprehensive evaluation of checks and balances institutions, and our results can serve as a normative benchmark to assess any such institution, regardless of its specific design. We illustrate the applicability of our normative benchmark within the context of constitutional review, a crucial pillar of established democracies, and, increasingly, of developing democracies.

Keywords: Institutional design; checks and balances; veto power; constitutional review.

* We thank Doug Bernheim, Josh Cohen, Jon Eguia, Jim Fearon, John Ferejohn, Sean Gailmard, Navin Kartik, Keith Krehbiel, Matt Jackson, Terry Moe, Mattias Polborn, Richard Van

Online Appendix available from:
http://dx.doi.org/10.1561/100.00013022_app
MS submitted 26 February 2013; *final version received 7 January 2014*
ISSN 1554-0626; DOI 10.1561/100.00013022

Essential to the Madisonian conception of checks and balances is the division of policy-making authority among multiple actors such that any one of them can veto proposed changes in policy. Thus, for example, the President can veto statutes passed by the Senate and House of Representatives; Congress can overturn international treaties signed by the President; and the Supreme Court can review the constitutionality of legislation passed by Congress and approved by the President.

There is more than one way to structure how the actors in any particular checks and balances setting interact. Consider, for example, interactions between two veto players who must agree before a new policy can be adopted. Possible ways to structure their interactions include: allowing one actor to propose a policy and the other then to accept it or not; allowing one actor to send a message about its policy preferences after which the other actor proposes a policy and the first actor then accepts it or not; or allowing each of the two actors to propose a policy and then bargain back and forth. In fact, there are infinitely many ways to specify the institutional arrangement under which actors interact, solely by varying what the actors can and cannot do and when they can and cannot do it.

Scholars as prominent as Thomas Hobbes, Adam Smith, and Frederick Hayek have observed that the institutional arrangements designed to coordinate actors' interactions shape which outcomes emerge. Even when the actors' information and preferences remain constant, different institutions generate different equilibrium outcomes. Because some equilibrium outcomes will be judged better than others, it follows that the study of checks and balances institutions is the proper level of analysis, which leads to questions such as: do some checks and balances institutions lead to outcomes that encourage more transparent behavior than others, all else equal? Do some checks and balances institutions lead to outcomes that minimize bargaining costs, all else equal? Do some checks and balances institutions better coordinate the players' interactions so as to avoid outcomes that none would have chosen?

To answer questions of this sort, we undertake a mechanism design analysis (Myerson, 2008). Specifically, we use this approach to investigate which, amongst all possible checks and balances institutions, meet

Weelden, seminar participants at Harvard University, Indiana University, New York University, Stanford University, University of Illinois, two anonymous reviewers and the editors of this journal for helpful comments and suggestions.

two desirable properties: strategy-proofness and Pareto efficiency. Strategy-proofness requires that a player lack incentives to act strategically and misrepresent her preference, regardless of what other players do. The Pareto efficiency criterion requires that the chosen outcome be such that no other feasible outcome is preferred by one player and not less preferred by others. Pareto efficient institutions avoid coordination failures and also reduce bargaining costs, all else equal, because there is no room jointly to bargain further to improve players' utilities.

Checks and balances institutions that meet the strategy-proofness and efficiency criteria, we show, are those in which the most preferred policy of the moderate player is the unique equilibrium outcome, where the moderate player is defined as the player preferring the less aggressive change from the policy status-quo, in situations when both players prefer changing the existing policy in the same direction. Moreover, such institutions are also the only checks and balances institutions that are strategy-proof and responsive.[1]

Our results can serve as a normative benchmark to assess checks and balances institutions. Manin *et al.* (1999) note the absence of a systematic way to compare checks and balances institutions. Our analysis alleviates this problem in that it facilitates a comprehensive evaluation of any checks and balances institution, regardless of its specific design. For example, one can take any institutional setting where two or more players need to agree to effect policy change, formalize that institution as a game, and then analyze its equilibrium outcomes. If the equilibrium outcome is the most preferred policy of the moderate player when players agree on the direction of policy change, then the institution in question is strategy-proof and efficient. If not, then the respective institution is either manipulable or inefficient (or both).

We show the applicability of our normative benchmark within the context of constitutional review. We focus specifically on the difference in timing of constitutional review as practiced in the United States and Europe to illustrate how our analysis facilitates normative assessments of checks and balances institutions. In the United States, constitutional review is mostly confined to specific controversies following governmental implementation of a policy, while in Europe constitutional review can also occur prior to the

[1] Responsiveness requires that the policy outcome be positively responsive to changes in the preferences of the players whenever both agents prefer a change from the status-quo policy in the same direction.

government's implementation of a policy (Ferejohn and Pasquino, 2002). We abstract and analyze the two approaches to constitutional review as simple games and show that the moderate's most preferred policy is the unique equilibrium outcome under the latter and not the former type of review. These results have implications for scholarly debates about how to structure constitutional review.

Our analysis contributes to two theoretical literatures. First, it contributes to a political economy literature on institutional design (for example, Buchanan and Tullock, 1962; Shepsle, 1979; Riker, 1982; Calvert, 1995; Ostrom, 2005). In the spirit of these studies, our work takes institutions as the proper level of analysis and shows some novel results about checks and balances institutions. Related research focuses on the conditions under which citizens benefit from dividing policy-making authority among multiple actors (Brennan and Hamlin, 2000; Acemoglu *et al.*, 2013) or assesses how checks and balances institutions (or multiple veto players) affect policy stability (Tsebelis, 2002), economic development (Cox and McCubbins, 2001), and policy credibility (North and Weingast, 1989; Keefer and Stasavage, 2003). We contribute to this literature by undertaking a comprehensive evaluation of any checks and balances institution and by providing a normative benchmark to assess any such institution, regardless of its specific design.

Second, our analysis contributes to a diverse and extensive political economy literature on veto bargaining (for example, Romer and Rosenthal, 1978; Matthews, 1989; Lupia, 1992; Persson *et al.*, 1997; McCarty, 1997; Diermeier and Myerson, 1999; Cameron, 2000; Bueno de Mesquita and Stephenson, 2007; Fox and Van Weelden, 2010; Gailmard and Hammond, 2011).[2] Most of this research analyzes the interactions of the players and the resulting equilibrium outcomes within specific institutional arrangements in which multiple players need to agree to effect policy change. Our mechanism design approach complements such game theoretic studies by showing results that apply to all possible institutional arrangements that could structure players' interactions.

1 The Model

Our formal analysis applies generally to any two (or more) interacting parties. For simplicity of exposition, we examine the interaction between two

[2] The above citations represent only some of the works in this area.

players, C and G.[3] Both players have preferences over a one-dimensional policy space, $X = \mathbf{R}$. An exogenous status-quo, q, exists. It can be viewed as the policy resulting from inaction. Without loss of generality, we normalize $q = 0$.

C's preference is represented by a continuous and single-peaked (about an ideal position c) utility function $U_C(\cdot, c)$ that satisfies the single-crossing property. G's preference over policies is represented by a continuous and single-peaked (about an ideal position g) utility function $U_G(\cdot, g)$ that satisfies the single-crossing property.[4] Single crossing is an appealing property, especially in the context of ideological preferences. Intuitively, it states that, given any two policies, if a player i prefers the more right-wing policy, then all players more right-wing than i (with ideal policies to the right of i's ideal policy) will also prefer the more right-wing of the two policies (Ashworth and Bueno de Mesquita, 2007).

C's ideal policy c and G's ideal policy g are private information, and we can think of each player as having a variety of possible types regarding her most preferred policy. Let the type of player C be given by $c \in \mathbf{R}$ and the type of player G by $g \in \mathbf{R}$. The players share a common prior about the joint distribution of types, which is given by $F(c, g)$ with full support on $\mathbf{R} \times \mathbf{R}$.[5]

An institution specifies the rules of the game under which players C and G interact. Our goal is to analyze comprehensively all possible institutions in which both players must agree to effect policy change in accordance with selected criteria. Although this task might appear intractable, the revelation principle facilitates it. For any given institution, let $\mathbf{s}^*$ be a Bayesian equilibrium of the game defined by the institution in question. The revelation principle states that, for any such Bayesian equilibrium, there exists a truth-revealing direct mechanism[6] that is payoff-equivalent with such Bayesian equilibrium $\mathbf{s}^*$ (Myerson, 1979). To analyze the set of all possible institutions, therefore, it is sufficient to consider only truth-revealing direct mechanisms. In a truth-revealing direct mechanism, the (message) action spaces

---

3 In the Appendix, we generalize our analysis to situations in which there are more than two players.

4 Formally, $U_I(x, i)$ satisfies the single-crossing property if for any $x_1 > x_2$ and $i_1 > i_2$, $U_I(x_1, i_2) \geq U_I(x_2, i_2)$ implies $U_I(x_1, i_1) > U_I(x_2, i_1)$, and $U_I(x_2, i_1) \geq U_I(x_1, i_1)$ implies $U_I(x_2, i_2) > U_I(x_1, i_2)$, where $I \in \{C, G\}$ and $i \in \{c, g\}$.

5 Full support is necessary here because we are interested in analyzing checks and balances institutions that satisfy selected criteria for all possible configurations of players' ideal policies.

6 A direct mechanism is a mechanism in which the action set equals the type set.

are precisely the type spaces, and in equilibrium all individuals reveal their true types.

One way of thinking about this is that instead of analyzing all possible institutional arrangements, we need only to study a simple setting in which the players' actions are to report their types and an outcome results as a function of the players' reported types. Thus a mechanism $x(c, g)$ specifies an outcome $x \in X$ as a function of C's and G's reported types. If such an outcome $x(c, g)$ satisfies certain properties, then all checks and balances institutions in which $x(c, g)$ is the unique equilibrium outcome satisfy those properties as well. As a result, the mechanism design approach allows us to establish general results about properties of checks and balances institutions.

Because we are interested in analyzing the properties of institutions in which both players must agree to effect policy change, this veto requirement implies that each player's utility from the outcome resulting under the respective institution must be at least what it would be if the status-quo policy is maintained. Therefore, the mechanism $x(c, g)$ must satisfy the following individual rationality constraints, which we label as the *checks condition*:

Checks: a mechanism $x(c, g)$ satisfies the checks condition if and only if $U_G(x(c, g), g) \geq U_G(0, g)$ and $U_C(x(c, g), c) \geq U_C(0, c)$ for any type profile (c, g).[7]

We refer to institutions that satisfy this checks condition as checks and balances institutions. We seek to characterize checks and balances institutions that meet two criteria: strategy-proofness and Pareto efficiency.

Strategy-proofness: a mechanism $x(c, g)$ is strategy-proof if and only if $U_C(x(c, g), c) \geq U_C(x(\widetilde{c}, g), c)$ and $U_G(x(c, g), g) \geq U_G(x(c, \widetilde{g}), g)$ for all c, $\widetilde{c}$, g, $\widetilde{g}$.

The strategy-proofness condition on $x(c, g)$ requires that truthful revelation is a dominant strategy for all players. Thus $x(c, g)$ is an incentive-compatible direct mechanism, and by the revelation principle, the set of all payoff pairs that can result from such mechanisms contains all the equilibrium payoff pairs of any process by which two players interact to determine

[7] These ex-post individual rationality constraints ensure that institution is robust in the sense that it does not depend on players' conjectures about the other players' behaviors (Chung and Ely, 2006).

a policy. Processes in this broader class may involve any well-specified combination of talk, voting, or other actions that together determine a policy outcome; their equilibria may involve all sorts of strategic manipulation, misdirection, or misrepresentation by players. Hence, we can, without loss of generality, abstract away all such details of the players' strategies and interactions, and focus on the strategy-proof direct mechanisms.

From a substantive perspective, institutions that are strategy-proof are relatively more transparent in determining how each player affects the outcome. If an institution is manipulable, there is room for players to act strategically within it. Such manipulation, when it occurs, makes it difficult to ascertain what players truly preferred. In contrast, in a strategy-proof institution, players do not benefit from manipulation because they have a dominant strategy to reveal their true preferences regardless of what others do. When there is no discrepancy between the players' true and stated preferences, it is easier to ascertain their true preferences. These transparency properties are particularly important in the context of representative institutions, where citizens need to assess their elected politicians' performances. Indeed at the constitutional convention of 1787, the Federalists and the Antifederalists debated at length about how to structure governmental institutions to make them accessible to ordinary citizens (Manin, 1994).

Pareto Efficiency: a mechanism $x(c, g)$ is Pareto efficient if and only if its outcome is Pareto efficient. Formally, for any (c, g), there is not another outcome x' such that $U_C(x', c) \geq U_C(x(c, g), c)$ and $U_G(x', g) \geq U_G(x(c, g), g)$, and at least one of the above inequalities holds with strict inequality.

The Pareto efficiency criterion requires that the chosen outcome is such that no other feasible outcome is preferred by one player, and not less preferred by the other. In our context, efficiency of $x(c, g)$ is equivalent to saying that for any c and g, $x(c, g)$ lies in the closed interval between c and g.

Pareto efficiency is a desirable property of checks and balances institutions for several reasons. First, it requires that the structure of interactions under a particular institution allow players to choose their own actions while avoiding outcomes that none would have chosen. In institutions where the interactions of two or more actors lead to a result that is not efficient, the outcomes are coordination failures. Efficiency thus implies the absence of coordination failures in checks and balances institutions.

Second, if an institution induces an outcome that is efficient, then the outcome is renegotiation proof and there can be no further improvements via bargaining in that institution. This stands in contrast with an institution in which the outcome is not renegotiation proof and where some players have an incentive to continue the bargaining process. As a result, bargaining costs are lower in an institution in which the outcome is efficient than in an institution in which the outcome is not, all else equal.

Third, if an institution induces an outcome that is efficient, and thus renegotiation proof, it will be stable in that the players lack incentives to coordinate on a new policy. The importance of policy stability as a property of institutions was noted in the Federalist papers. For example, in Federalist No. 62, Hamilton writes "The internal effects of mutable policy are still more calamitous. It poisons the blessing of liberty itself. It will be of little avail to people, that the laws are made by men of their choices, if ...they can be repealed or revised before they are promulgated, or undergo such incessant change that no man who knows what the law is today, can guess what it will be tomorrow." The rule of law literature also underlines the importance of policy stability as a precondition to predict what actions are or are not allowed under the law (Fuller, 1969). From this perspective, policy stability enables individuals to know in advance the range of activities in which they are free to do as they please without being exposed to government coercion (Hayek, 1960).

We also investigate the set of checks and balances rules that are strategy-proof and responsive. That is, in addition to strategy-proofness, we impose responsiveness rather than efficiency as a desirable property in our characterization. We show that the set of checks and balances institutions that satisfies strategy-proofness and efficiency and the set that satisfies strategy-proofness and responsiveness are the same.

Responsiveness: a mechanism $x(c, g)$ is (strictly) responsive to players' preferences if and only if for any $c' > c \geq 0$ and any $g' > g \geq 0$, $x(c', g') > x(c, g)$, and for any $c' < c \leq 0$ and any $g' < g \leq 0$, $x(c', g') < x(c, g)$.

Responsiveness requires that the outcome $x(c, g)$ respond positively to changes in the players' preferences when both players want a change from the policy status-quo, 0. The responsiveness criterion requires no more than absence of a situation where both agents want a policy change in the same direction and the policy does not change (Riker, 1982).

We do not propose that these criteria are the only desirable properties of checks and balances institutions. Rather, we view our analysis as a first step in evaluating checks and balances institutions more generally.

Someone seeking to consider the effects of additional properties can assess, for example, whether any checks and balances institution within the set of strategy-proof and efficient institutions satisfies the additional properties. A negative answer implies a trade-off among different, desirable criteria. This knowledge can help scholars and practitioners to make informed assessments regarding which properties are more suitable for the institutional setting under investigation.

In the subsequent analysis, we use the following definition for the moderate player's most preferred outcome, $m(c, g)$:

Definition. *For any c and g with the same signs, the outcome $m(c, g)$ is defined as the most preferred policy of player whose ideal policy is closer to the status quo $q = 0$; that is, $m(c, g) = \mathbf{min}\{c, g\}$ if $c, g \geq 0$ and $m(c, g) = \mathbf{max}\{c, g\}$ if $c, g \leq 0$.*

2 The Analysis

When players disagree about the direction of policy change (i.e., $c < 0 < g$ or $g < 0 < c$), the policy outcome of any checks and balances institutions does not depend on the rules of the game under which players interact. In such situations, the resulting outcome is the status-quo policy, because it is the only outcome that satisfies the checks condition.[8] As such, the analysis regarding which mechanisms satisfy the checks, strategy-proofness, and efficiency conditions is simple: the unique mechanism that satisfies these conditions is $x(c, g) = 0$ when $c < 0 < g$ or $g < 0 < c$.[9]

When players agree on the direction of policy change (i.e., $c, g \geq 0$ or $c, g \leq 0$), the resulting outcome depends on the institutional arrangement under which players interact. In such situations, the strategic interaction between G and C has elements of both common interest and conflict. The common interest stems from both players preferring a change from the status-quo policy. The conflict stems from the players disagreeing about the policy that should replace the status-quo. Such disagreement may give each player an incentive to manipulate the bargaining process to induce a

---

[8] Any other policy outcome $p \neq 0$ will make one player worse off than the status-quo policy and thus the respective player will oppose such policy change.

[9] Note that, when $c < 0 < g$ or $g < 0 < c$ the status-quo policy is Pareto efficient and also strategy-proof since players cannot gain by misreporting their preferences.

policy change closer to its most preferred policy. The specific institution under which the players interact is therefore crucial for the policy outcome; it determines how each player reveals its private information and whether the players do or do not successfully coordinate their actions to change the status-quo.

Because only when players agree on the direction of policy change does the resulting outcome depend on the institutional arrangement under which players interact, for simplicity of exposition, we focus our analysis on this scenario. Specifically, we analyze the case in which $c, g \geq 0$ since $c, g \leq 0$ is similar. We re-state the formal results for the general case in which $c \in \mathbf{R}$ and $g \in \mathbf{R}$ at the end of this section.

We first characterize some properties of strategy-proof mechanisms.[10] For strategy-proof mechanisms, players' strategies must be optimal for each type, independent of what other players of any type do. Because each player type is required to have a dominant strategy, strategy-proof mechanisms produce the desired outcomes independent of what players think about each other.

Mechanisms that are strategy-proof for C, given an arbitrary type of G, and mechanisms that are strategy-proof for G, given an arbitrary type of C, must have the following properties:

Lemma 1 *For any g, any mechanism $x(c, g)$ that is strategy-proof for player C is weakly increasing in c; if $x(c, g)$ is strictly increasing in c on an open interval (c_1, c_2), then $x(c, g) = c$ on (c_1, c_2). For any c, any mechanism*

---

[10] Moulin (1980) and Barbera and Jackson (1994) characterize strategy-proof social choice functions on the domain of all single-peaked preferences. Our purpose here is to analyze checks and balances institutions on the basis of selected criteria; nevertheless our work also differs from existing work on strategy-proof mechanisms. First, the properties of strategy-proof mechanisms we characterize hold for mechanisms that are strategy-proof on the domain of any specific single-peaked single-crossing $(U_C(\cdot, c), U_G(\cdot, g))$. Thus there are infinitely many strategy-proof mechanisms that are not strategy-proof on the domain of all single peaked preferences. For example, given any single-peaked single-crossing (U_C, U_G), for any $a > 0$ define $\bar{c}_a$ by $U_C(0, \bar{c}_a) = U_C(a, \bar{c}_a)$ and $\bar{g}_a$ by $U_G(0, \bar{g}_a) = U_G(a, \bar{g}_a)$, the mechanism $x(c, g) = 0$ if $c \in [0, \bar{c}_a]$ or if $g \in [0, \bar{g}_a]$, $x(c, g) = a$ if $c \in (\bar{c}_a, a]$ and $g > \bar{g}_a$, or if $c > \bar{c}_a$ and $g \in (\bar{g}_a, a]$, and $x(c, g) = \mathbf{min}\{c, g\}$ otherwise, is strategy-proof on the domain of (U_C, U_G), but not on the domain of all single-peaked preferences. Second, we work on the domain of single-crossing single-peaked preference functions and the class of strategy-proof mechanisms can only be larger on this domain. This is the case because strategies that were dominant remain dominant while strategies that were not dominant in the larger domain (single-peaked) can become dominant on the smaller domain (single-crossing single-peaked).

$x(c,g)$ that is strategy-proof for player G is weakly increasing in g; if $x(c,g)$ is strictly increasing in g on an open interval (g_1,g_2), $x(c,g)=g$ on (g_1,g_2).

Proof: See online appendix. ■

This result is well known in the mechanism design literature (Melumad and Shibano, 1991). For example, it suggests that strategy-proof mechanisms consist either of regions where the outcome is the same regardless of player C's type or of regions where the outcome is C's ideal policy. To see the intuition for why the mechanism has to be $x(c)=c$ when the mechanism is not flat, suppose instead that the mechanism is $x(c)=c+\epsilon$. Then, there will be some types of player C that will have an incentive to report $c-\epsilon$ instead of c in order to ensure a better outcome; and, therefore, the strategy-proofness condition will not be met.

Thus strategy-proof mechanisms implement either the player's ideal policy or a constant policy, or combinations of player's ideal policies and constant policies. For example, the mechanisms $x(c,g)=0$ and $x(c,g)=\mathbf{min}\{c,g\}$ are both strategy-proof. The second mechanism is strategy-proof because, for each player, the outcome is either its own ideal policy or some policy lower than its ideal policy. In the former case, a player has no incentive to deviate and, in the latter case, the only way to change the outcome is to announce and implement an even lower policy, which would make the player worse off. The first mechanism is trivially strategy-proof; the outcome is the status quo, $q=0$, regardless of what the players are doing, and therefore the players do not have an incentive to misreport their preferences. Note that strategy-proof mechanisms can be complicated in the sense that, in some intervals, a player's ideal policy is implemented and in other intervals a constant policy is implemented.[11]

To show what each of the three conditions-checks, strategy-proofness, and efficiency-requires, we offer examples of mechanisms that violate one of the conditions. First, the mechanism $x(c,g)=g$ for any c,g is strategy-proof and efficient but violates the checks conditions. The mechanism is strategy-proof because players cannot become better off by misrepresenting their preferences. The mechanism is also efficient because the outcome is one

[11] For example, given any single-peaked single-crossing (U_C,U_G), define $\bar{c}$ by $U_C(0,\bar{c})=U_C(2,\bar{c})$ and $\bar{g}$ by $U_G(0,\bar{g})=U_G(2,\bar{g})$, the following mechanism is strategy-proof when $c,g\geq 0$: $x(c,g)=0$ if $c\in[0,\bar{c}]$ or if $g\in[0,\bar{g}]$, $x(c,g)=2$ if $c\in(\bar{c},2]$ and $g>\bar{g}$, or if $c>\bar{c}$ and $g\in(\bar{g},2]$, and $x(c,g)=\mathbf{min}\{c,g\}$ otherwise.

of the players' ideal policies. However, the mechanism violates the checks condition because there will be situations in which a player will be better off with the status-quo policy. For example, let $c = 1$ and $g = 10$. For any given symmetric preference, the player with ideal policy at $c = 1$ is better off with the status quo policy 0 than with the outcome generated by mechanism $x(c, g) = 10$.

Second, for any symmetric preference,[12] the checks and balances mechanism $x(c, g) = \mathbf{min}\{2\mathbf{min}\{c, g\}, c\}$ for $c, g \geq 0$ satisfies efficiency and not strategy-proofness.[13] This mechanism is efficient because $\mathbf{min}\{c, g\} \leq \mathbf{min}\{2\mathbf{min}\{c, g\}, c\} \leq \mathbf{max}\{c, g\}$ for $c, g \geq 0$. However, it is not strategy-proof, as the following counter example illustrates. Let $g = 2$ and $c = 5$. Then the outcome is $x(c, g) = 4$. But G has an incentive to deviate and announce $g' = 1$ because the outcome then is $x(c, g') = 2$, which makes G better off.

Third, the checks and balances mechanism $x(c, g) = 0$ for all c and g satisfies strategy-proofness and not efficiency. We noted earlier that this mechanism is trivially strategy-proof because it does not depend on the announced values of the players' ideal points. Moreover, the outcome is always the status quo. The mechanism is not efficient because, for $c > 0$ and $g > 0$, the outcome $q = 0$ is not efficient.

A unique mechanism satisfies the checks, strategy-proofness, and efficiency conditions: the players report their ideal policies and the outcome consists of implementing the ideal policy of the player closer to the status-quo policy, $m(c, g)$. The result is as follows:

Proposition 1 *The unique mechanism that satisfies the checks, strategy-proofness, and Pareto efficiency conditions is* $x(c, g) = m(c, g)$ *for* $c, g \geq 0$.

Proof: See online appendix. ■

Proposition 1 can serve as a benchmark to assess designs of existing checks and balances institutions, including the interactions between two chambers of bicameral legislatures, between legislatures and executives, and between courts and governments. It suggests that if the equilibrium outcome in any checks and balances institution is not the moderate player's ideal policy,

---

[12] All symmetric single-peaked functions satisfy the single-crossing property.

[13] The mechanism satisfies the checks condition because $|\mathbf{min}\{2\mathbf{min}\{c, g\}, c\} - c| \leq c$ and $|\mathbf{min}\{2\mathbf{min}\{c, g\}, c\} - g| \leq g$ for $c, g \geq 0$.

$m(c, g)$ when both players want a policy change in the same direction, then the institution in question is either manipulable or inefficient (or both).

We also investigate the set of mechanisms that satisfies the checks, strategy-proofness, and responsiveness conditions. That is, along with strategy-proofness, we impose responsiveness rather than Pareto efficiency as desirable properties in our characterization.

We offer examples of mechanisms that satisfy the checks condition but violate either strategy-proofness or responsiveness to show what each property requires. First, for any symmetric preference, the checks and balances mechanism $x(c, g) = \mathbf{min}\{2\mathbf{min}\{c, g\}, \frac{c+g}{2}\}$ for $c, g \geq 0$ satisfies responsiveness and not strategy-proofness.[14] This mechanism is responsive because for any $c' > c \geq 0$ and $g' > g \geq 0$, $\mathbf{min}\{c', g'\} > \mathbf{min}\{c, g\}$ and $\frac{c'+g'}{2} > \frac{c+g}{2}$. Therefore, $x(c', g') = \mathbf{min}\{2\mathbf{min}\{c', g'\}, \frac{c'+g'}{2}\} > x(c, g) = \mathbf{min}\{2\mathbf{min}\{c, g\}, \frac{c+g}{2}\}$. However, the mechanism is not strategy-proof, as the following counter example illustrates. Let $g = 100$ and $c = 120$. Then, the outcome is $\frac{c+g}{2} = 110$. But C has an incentive to deviate and announce $c' = 140$ because the outcome is $\frac{c'+g}{2} = 120$ and C is better off.

Second, the checks and balances mechanism $x(c, g) = \mathbf{min}\{a, c, g\}$, where $a > 0$ is any positive constant, satisfies strategy-proofness and not responsiveness. This mechanism is strategy-proof because, for any player, the outcome is either its ideal policy or is lower than its ideal policy when telling the truth. If a player misrepresents its ideal point, the player can only change the outcome to a lower policy outcome, which makes it worse off. This mechanism is not responsive because for any $c' > c > a$ and $g' > g > a$, $x(c', g') = x(c, g) = a$.

We have the following result:

Proposition 2 *The unique mechanism that satisfies the checks, strategy-proofness, and responsiveness conditions is* $x(c, g) = m(c, g)$ *for* $c, g \geq 0$.

Proof: See online appendix. ■

The set of checks and balances institutions that are strategy-proof and responsive and those that are strategy-proof and efficient are the same. Proposition 2 thus suggests that a checks and balances institution in which

---

[14] The mechanism satisfies the checks condition because $|\mathbf{min}\{2\mathbf{min}\{c, g\}, \frac{c+g}{2}\} - c| \leq c$ and $|\mathbf{min}\{2\mathbf{min}\{c, g\}, \frac{c+g}{2}\} - g| \leq g$ for $c, g \geq 0$.

the outcome is not the moderate player's ideal policy, $m(c, g)$ is either manipulable or not responsive (or both).

Note that the mechanism $x(c, g) = m(c, g)$ is also anonymous, although we did not use anonymity in our characterization. Anonymity is a form of equity in that it requires the mechanism $x(c, g)$ not to depend on which player has which specific preference, so that interchanging the players' preferences will result in the same outcome.[15] Thus the mechanism $x(c, g) = m(c, g)$ is the only mechanism that satisfies strategy-proofness, responsiveness, efficiency, and anonymity within checks and balances institutions, when players agree on the direction of policy change.[16]

We focus the previous analysis on situations in which $c, g \geq 0$ (or $c, g \leq 0$) because, when players disagree about the direction of policy change, the resulting policy outcome is the status-quo, regardless of the institutional arrangement under which players interact. We can restate our formal results for the general situation in which $c \in \mathbf{R}$ and $g \in \mathbf{R}$. Proposition 1 can be restated as follows:

Proposition 1′. *When players agree on the direction of policy change (i.e., $c, g \geq 0$ or $c, g \leq 0$), the unique mechanism that satisfies the checks, strategy-proof, and Pareto efficiency conditions is $x(c, g) = m(c, g)$. When players disagree on the direction of policy change (i.e., $c < 0 < g$ or $g < 0 < c$), the unique mechanism that satisfies the checks, strategy-proof, and Pareto efficiency conditions is $x(c, g) = 0$.*

Proof: See online appendix. ∎

Proposition 2 can be restated similarly for $c \in \mathbf{R}$ and $g \in \mathbf{R}$.

Our analysis can be extended to a situation in which there are $N > 2$ decision-makers who must agree on changing an existing policy. We provide this analysis in the Online Appendix.

Our results allow us to make normative evaluations of checks and balances institutions. In principle, we can take any institutional setting where two or

[15] Formally, the mechanism $x(c, g)$ is anonymous if $x(c, g) = x(g, c)$.

[16] Although we worked on the domain of continuous single-crossing single-peaked preferences, for continuous single-peaked utilities that do not satisfy the single-crossing property, the mechanism $x(c, g) = m(c, g)$ satisfies both strategy-proofness and efficiency or strategy-proofness and responsiveness. Thus, the mechanism $x(c, g) = m(c, g)$ is the unique mechanism that satisfies strategy-proofness and responsiveness or strategy-proofness and efficiency conditions for all continuous and single-peaked utilities within the context of checks and balances institutions.

more players need to agree to effect policy change, formalize that institution as a game, and then analyze its equilibria outcomes. If the unique equilibrium is $x(c, g) = m(c, g)$ when both players agree on the direction of policy change, then that institution is strategy-proof and responsive, and strategy-proof and efficient. In contrast, if the equilibrium outcome is not $x(c, g) = m(c, g)$, the institution in question is either manipulable or inefficient (or both), and so on.

To illustrate how our analysis can be used to make normative assessments of checks and balances institutions we apply it within the context of constitutional review.

3 Application to Constitutional Review

Liberal societies share a normative commitment to constitutional review, as most recently evidenced by its widespread adoption after the Second World War. Constitutional review, in turn, means that governments cannot trample on constitutionally-protected rights, including even when they enjoy the support of a majority of citizens; and that the constitutional court holds the authority to make sure they do not.

Scholars routinely distinguish the American from the European practice of constitutional review (Ferejohn and Pasquino, 2002). The two practices differ in several ways but an especially notable difference is the timing of review. Although there are exceptions, constitutional review in the United States is mostly limited to legal suits claiming infringements of protected rights that individuals bring following the government's implementation of a policy.[17] In Europe constitutional review can additionally occur prior to governmental implementation of a policy (Ferejohn and Pasquino, 2004).

Focusing on the difference in timing, we abstract and analyze the two approaches to constitutional review as simple games in order to assess their equilibria on the basis of our normative benchmark. Let the government's preference be represented by a twice continuously differentiable, single crossing, and single-peaked (about an ideal position g) utility function $U_G(\cdot)$ and the court's preference be represented by a a twice continuously differentiable, single crossing, and single-peaked (about an ideal position c) utility

[17] The court sometimes issues preliminary injunctions and declaratory judgements, mostly in the context of First Amendment rights.

function $U_C(\cdot)$. The government's and court's policy preferences are private information; the government type is given by $g \in \mathbf{R}_+$ and the court type by $c \in \mathbf{R}_+$. The players share a common prior about the joint distribution of types given by $F(c, g)$ with full support on $\mathbf{R}_+ \times \mathbf{R}_+$.[18]

The government's strategy is to choose a policy $p \in \mathbf{R}_+$. The court's strategy is to choose a legal limit on the government's policies $\ell \in \mathbf{R}_+$, with all policies $p \leq \ell$ being legal and policies $p > \ell$ being illegal. The court's strategy is consistent with the scholarly literature noting that the court's influence on policy typically consists of placing bounds on what the government can and cannot do in specific situations (Kornhauser, 1992; Barak, 2006; Fox and Stephenson, in press). This is the case because, by the very nature of the judicial process, the court's influence over policy mostly consists of answering (binary) questions of law (Kornhauser, 1992)[19] and because the courts do not implement policies.[20]

The outcome of the game is the status-quo $x = 0$ if $p > \ell$ and $x = p$ if $p \leq \ell$.[21] The difference between the two institutions consists in the timing of the interaction. In one institution, ex-ante review, the court choses the legal limit ℓ and then the government implements a policy p. In the other institution, ex-post review, the government implements a policy p first and then the court defines the range of legal policies.[22]

We first characterize the equilibrium of the ex-ante institution. The unique equilibrium outcome in this institution is the same as the mechanism defined

[18] Similar to the main analysis, we focus on the case in which $g \in \mathbf{R}_+$ and $c \in \mathbf{R}_+$ to see whether the equilibrium under each of these two institutions is $m(c, g)$.

[19] For example: do suspected terrorists have a constitutional right to habeas corpus? Are certain enhanced interrogation techniques legal in the context of terrorism prevention? In providing answers to such questions, the court influences policy by putting a bound on what the government can and cannot do.

[20] Suppose that the legal question is how many days the government can keep a suspected terrorist in detention, without access to legal representation. Suppose the court's most preferred policy is five days. The court can state that it is legal to keep a suspected terrorist for five days without legal rights; however, because the court does not implement this policy, such a ruling can only bound what the government can and cannot do. The government can allow the right to legal representation after only three days if it so wishes.

[21] Note that this specification implies that the government complies with the court decision when the court declares a policy illegal. For analyses of how legal limits and judicial rulings can be self-enforcing, see Weingast (1997) and Dragu and Polborn (2013).

[22] Note that we could introduce an additional stage in the ex-ante institution: the government sends a policy proposal y for the court to review before the court chooses the legal limit ℓ. Adding such an additional stage does not affect the equilibrium analysis. The important institutional aspect is whether the court can review the legality of policy before or after policy implementation.

in Proposition 1. The strategy profile $\ell(c) = c$ and $p(\ell, g) = \mathbf{min}\{g, \ell\}$ is an equilibrium strategy, resulting in the unique perfect Bayesian Nash equilibrium outcome $x(c, g) = m(c, g)$.

We prove this result by backward induction. In the second stage, the government's belief about the court's type after each observed limit is irrelevant because it does not affect the government's payoff. That is, if the legal limit chosen by the court in the first stage is ℓ for any court's strategy and any government's beliefs, then $p(\ell, g) = \mathbf{min}\{g, \ell\}$ is the unique optimal strategy for the government. In the first stage, if the court deviates to $\ell' < c$, then for all $g \leq \ell'$, the court receives the same payoff; for all $g \in (\ell', c)$, the court is worse off because the outcome is g if it chooses c and $\ell' < g \leq c$ if it chooses ℓ'; and for all $g > c$, the court is worse off because the outcome is c if it chooses c and $\ell' < c$ if it chooses ℓ'. A similar argument shows that the court also has no incentive to deviate to $\ell' > c$. Thus the following proposition:

Proposition 3 *The unique equilibrium outcome of the ex-ante institution is $x(c, g) = m(c, g)$.*

Proof: In text. ■

More generally, Proposition 3 suggests a general institutional arrangement through which one can obtain the mechanism $x(c, g) = m(c, g)$ if $c, g \geq 0$ as the unique equilibrium outcome. The moderate player's ideal policy is the equilibrium outcome in the institution where C first defines a legal limit $\ell \in \mathbf{R}_+$ and G then chooses a policy $p \in \mathbf{R}_+$; the outcome is $x = p$ if $p \leq \ell$ and $x = q$ if $p > \ell$.

We also solve for the equilibrium of the ex-post institution by backward induction. Here the court's belief about the government's type after each observed policy is irrelevant because it does not affect the court's payoff. Thus the court's optimal choice after any p chosen by the government is the following: a court type c chooses the legal limit $\ell = c'$ and thus all policies $p \leq c'$ are legal while policies $p > c'$ are illegal, where c' is such that a court type c is indifferent between the status-quo and the policy c', i.e., $U_C(0, c) = U_C(c', c)$.

Note that there is a range of policies, $p \in [c, \mathbf{min}\{c', g\}]$, that the government prefers over c when $c < g$ and that the court will find legal ex-post. Thus it might pay for the government to implement a policy before the court reviews its legality since the court might accept a policy closer to what the government prefers. However, the government does not know c, and thus takes a gamble: although it might get away with a policy $p > c'$, the resulting outcome will be the status-quo if the court finds $p > c'$ illegal. Given the court's equilibrium strategy, the government will never choose any policy higher than g, and thus the government's maximization problem is the following:

$$\mathbf{argmax}_{p \in [0,g]} \{\Pr(c' \geq p) \cdot U_G(p,g) + \Pr(c' < p) \cdot U_G(0,g)\}$$

in which $\Pr(c' \geq p)$ is the probability that the court type c finds p legal. The trade-off the government faces is the following. If the government implements a policy farther from the existing status-quo (and closer to its most preferred position), its payoff is bigger when the court upholds it. However, the government's choice also affects the probability that the policy is legal. A policy farther from the existing status-quo increases the probability of the court finding the policy illegal, meaning that such a choice could result in a lower payoff for the government. Note that for $c, g \geq 0$, there is a positive equilibrium probability, $\Pr(c' < p) > 0$, that the court rejects the government policy given that the distribution of c has full support. This result follows:

Proposition 4 *In any equilibrium of the ex-post institution, for any* c, g *such that* $\mathbf{min}\{c, g\} > 0$, *there is a positive probability that* $x(c, g) = 0$.

Proof: In text. ■

Proposition 4 suggests that the equilibrium under the ex-post institution cannot always be $x(c, g) = m(c, g)$ for $c, g \geq 0$ because there are situations in which the court vetoes the government's policy whenever both players prefer changing the status-quo.[23] Because the unique equilibrium outcome is not always the moderate player's most preferred policy, this institution can be manipulable, inefficient, or both.

Propositions 3 and 4 have some normative implications for the practice of constitutional review. First, they suggest that the ex-ante review institution is both strategy-proof and efficient, and strategy-proof and responsive, while

[23] The same conclusion holds if the court can send the government legal advice (a cheap talk message) before the government choses a policy, and the court can review the constitutionality of the policy only after implementation.

ex-post review is not necessarily so. Of course, specific institutional details of constitutional review might be the product of historical evolution or the short-run strategic choices of important political actors at the constitutional table. Although our analysis does not address these contingencies, it provides a normative justification for one institutional aspect of constitutional review: allowing court review prior to governmental implementation of a policy.

Second, James Thayer (1893) advanced an important criticism of constitutional review more than a century ago. Very possibly, he argued, constitutional review will cause members of elected branches of government to act strategically and allow unconstitutional policies they oppose on the assumption they know they can rely on the court to address such concerns (Tushnet, 1999). Our analysis suggests that Thayer's criticism likely applies to a specific institutional design, American (ex-post) constitutional review, rather than to constitutional review more generally. Since it is strategy-proof, ex-ante review does not create ''judicial overhang,'' implying that members of elected branches cannot benefit from acting strategically. Should the strategic behavior that might arise under ex-post judicial review reach the point of diluting elected officials' sense of constitutional responsibility, as Thayer feared, allowing ex-ante review as well can, in principle, overcome the problem.

Third, ex-post constitutional review can sometimes produce Pareto inefficient policies. Both the government and court might prefer a change from the policy status quo, but because the government implements a policy under uncertainty, the court might reject the policy on grounds that it fall outside the court's range of acceptability. When such policy coordination failure actually occurs, scholars sometimes direct their criticism to the particular agents involved in the policy process: the President or Congress were too aggressive in their policy choices; the Supreme Court justices were out of tune with the majority preferences of the day; and so forth. Although such criticisms might sometimes be warranted, Propositions 3 and 4 indicate that policy coordination failures could be caused by the design of constitutional review. Then, assigning responsibility to the agents acting under the particular institution can be misleading.

Our purpose in this section was to illustrate the applicability of our analysis to the study of one important checks and balances institution, not to make a comprehensive evaluation of the two systems of constitutional review.

We approached the comparison of the two institutions through the lens of our normative benchmark. This does not imply that there cannot be other criteria to compare the two constitutional systems, especially given the fact that the two institutions vary along many dimensions.

Consider, for example, an alternative perspective by which to evaluate the two institutions, namely the availability of information pertinent to the legality of policies. Some scholars argue that situations exist where seeing the policy in force can generate such information. Landes and Posner (1994) write that when deciding before rather than after the government implements a policy, a court sometimes lacks "the benefit of information generated by the act itself." And when the informational benefits from seeing the policy in action dominate other considerations, ex-post constitutional review can be more attractive for the court.[24]

Our normative assessment of the two institutions applies to situations in which the action of the government does not generate information relevant for its legality. For instance, Adler (1998) argues that addressing legal challenges in the realms of constitutional rights does not depend on the specifics of policies that might infringe upon those rights because constitutional rights are rights against rules, which implies that the legality of governmental action depends on answering questions of law. Consider, for example, the legal question of whether non-citizens who are suspected terrorists have a constitutional right not to be tortured. The government's action to torture such individuals does not generate any information relevant to this legal question. The court will not have more information pertinent to answering this legal question following government engagement in torture than prior to it.

Even in situations in which seeing the policy in force generates information useful to assess its legality, our analysis can help to understand trade-offs among various institutional arrangements. In such cases, our analysis suggests that there might be a trade-off between institutional designs that are efficient and non-manipulable and those that provide more information to assess the legality of policy. Such knowledge can help scholars and practitioners to make informed assessments regarding which constitutional review arrangement is most suitable in various situations.

---

[24] For such an analysis, see Rogers and Vanberg (2002).

4 Multidimensional Analysis

The previous analysis provides a normative benchmark to assess the equilibria of checks and balances institutions in the context of a unidimensional policy space. The unidimensional analysis is both empirically justifiable and theoretically relevant. From an empirical perspective, it is consistent with an extensive literature showing that, in a variety of legislative settings, a single dimension captures most of the variation in legislative voting. The seminal work is Poole and Rosenthal (1997), and numerous subsequent studies have found that roll call voting is low-dimensional in legislatures throughout the world (Londregan, 2000; Morgenstern, 2004; Hix *et al.*, 2006; Voeten, 2000, among others).

Our analysis also comports with theoretical research that builds upon the Madisonian notion of divided authority among multiple political actors, whereby such any one actor can veto proposed changes in policy. This game-theoretic work includes models of pivots and agenda-setting (Romer and Rosenthal, 1978; Lupia, 1992; Krehbiel, 1998; Primo, 2002; Cox and McCubbins, 2005, among others) and models of bargaining between the President and Congress (Matthews, 1989; McCarty, 1997; Cameron, 2000; Cameron and McCarthy, 2004, among others). These veto bargaining models vary considerably in terms of information structure, the players' strategies, and the timing of interactions. Mechanism design analysis is an important next-step in this theoretical literature because it facilitates the establishment of welfare properties of checks and balances institutions without assuming a specific bargaining protocol. Thus the results of mechanism design analysis can serve as normative benchmarks by which to assess the equilibria of any veto bargaining model, regardless of its specific institutional details. To achieve these benchmarks requires that the mechanism design analysis assume the same unidimensional policy space that existing game-theoretic bargaining models assume.

Although a unidimensional mechanism design analysis is both empirically justifiable and theoretically relevant, we nevertheless want to underline that the normative benchmark of the previous section, the moderate rule, is not an anomalous result driven solely by the dimensionality assumption. To this end, we investigate the properties of the moderate mechanism in a multidimensional policy space.

Two important theoretical findings shape the multidimensional analysis. First, non-dictatorial strategy-proof mechanisms do not exist on the general domain of non-separable multidimensional single-peaked preferences (Border and Jordan, 1983). To ensure that mechanisms satisfying the checks and strategy-proofness conditions exist, we confine our analysis to single-peaked preferences that are separable. Second, even on single-peaked preference domains on which strategy-proof mechanisms exist, there is a generic tension between strategy-proofness and Pareto efficiency in multiple dimensions (Border and Jordan, 1983; Kim and Roush, 1984). Therefore, we relax the efficiency requirement, as discussed below, to investigate the set of mechanisms that satisfy the checks and strategy-proofness conditions, as well as some weaker notions of efficiency.

Similar to the unidimensional analysis, we examine the interaction between two players, C and G.[25] Both players have preferences over an n-dimensional policy space, $X = \mathbf{R}_+^n$, and the status quo policy is $(0, \ldots, 0)$.[26] For each player $i \in \{C, G\}$, the preference over policy $(x_1, x_2, \ldots, x_n) \in \mathbf{R}_+^n$ is represented by utility of the form

$$u_i(x_1, x_2, \ldots, x_n) = -\sum_{j=1}^{n} \alpha_j (x_j - p_j^i)^2, \tag{1}$$

where $\alpha_j \geq 0$ for all j and $p^i = (p_1^i, p_2^i, \ldots, p_n^i) \in \mathbf{R}_+^n$ is player i's ideal policy. That is, the preference domain of each player is $\{u(x_1, x_2, \ldots, x_n) = -\sum_{j=1}^{n} \alpha_j (x_j - p_j)^2 | \alpha_j \geq 0, p_j \geq 0, \forall j\}$. Note that the Euclidean preference, i.e., a preference of the form $\{(x, y) : d(x, p^i) \leq d(y, p^i)\}$, where p^i is player i's ideal point and d is the Euclidean distance, is one example of such preferences; the Euclidean preference can be obtained by setting $\alpha_j = 1$ for all j.

Strategy-proof mechanisms exist when players' preferences are represented by Expression (1). On such preference domains, however, there is a tension between strategy-proofness and Pareto efficiency. To illustrate this tension, let the players' preferences be given by the Euclidean preference, and consider two possible generalizations of the unidimensional moderate mechanism to the multidimensional case.

First, consider the following issue-by-issue moderate mechanism:

$$\mathbf{m}(p^C, p^G) = (\mathbf{min}\{p_1^C, p_1^G\}, \mathbf{min}\{p_2^C, p_2^G\}, \ldots, \mathbf{min}\{p_n^C, p_n^G\}),$$

[25] The analysis applies generally to any two (or more) interacting parties.

[26] That is, we focus on the case in which each player prefers a change in the positive direction on each dimension, which is a generalization of our basic unidimensional analysis. The analysis can also be generalized to all other cases, as in the unidimensional analysis.

where $j = 1, \ldots, n$ represents the issue dimensions and p_j^i is player i's ideal policy on issue dimension j for $i \in \{C, G\}$. That is, if we consider the components of the players' ideal policies on each dimension, i.e., the pair (p_j^C, p_j^G), the policy outcome under this mechanism is to choose, on each dimension, the ideal policy of the player who is closer to the status quo, i.e., $\mathbf{min}\{p_j^C, p_j^G\}$.

The issue-by-issue moderate mechanism $\mathbf{m}(p^C, p^G)$ satisfies the strategy-proofness and checks conditions, although not the efficiency condition. The mechanism $\mathbf{m}(p^C, p^G)$ is strategy-proof because the only way a player can influence the outcome by lying is to induce an even lower policy on all dimensions, which does not benefit him/her. This mechanism satisfies the checks condition, because for any player $i \in \{C, G\}$, the resulting policy is such that $0 \leq \mathbf{m}(p^C, p^G) \leq p^i$, and by definition $u_i(\mathbf{m}(p^C, p^G)) = -\sqrt{\sum_{j=1}^n (\mathbf{m}(p^C, p^G)_j - p_j^i)^2} \geq u_i(0) = -\sqrt{\sum_{j=1}^n (0 - p_j^i)^2}$. To see that this mechanism is not efficient, consider the case of two dimensions, $n = 2$, and let player C's ideal policy be $(2, 1)$ and player G's ideal policy be $(1, 2)$. The outcome of this mechanism is $(1, 1)$, which is not efficient because there exists another policy, say $(\frac{3}{2}, \frac{3}{2})$, that improves the utilities of both the players.[27]

Second, consider the following moderate mechanism, conceptualized in terms of the preferred policy of the player whose ideal policy is closer to the status quo:

$$\mathbf{m}'(p^C, p^G) = p^{i^*},$$

where p^{i^*} is the ideal policy of player $i^* \equiv \mathbf{argmin}_{i \in \{C,G\}} d(0, p^i)$, with d representing the Euclidean distance function. That is, this mechanism chooses the preferred policy of the player whose ideal policy is closer to the status quo.

The mechanism $\mathbf{m}'(p^C, p^G)$ is efficient because it always implements the ideal policy of one of the players. However, it is not strategy-proof, nor does it satisfy the checks condition. Consider two dimensions, $n = 2$, and let player C's ideal policy be $(1, 0)$ and player G's ideal policy be $(0, 2)$. The outcome of this mechanism is $(1, 0)$. Now consider the incentives of player G.

[27] Given Euclidean preference, Pareto efficiency in multiple dimensions implies that the policy in question must lie on the straight-line connecting the ideal policies of the two players, and the mechanism in question is not efficient unless the two players' ideal policies on at least one of the two dimensions are the same.

If he/she reports his/her ideal policy truthfully, the outcome is $(1, 0)$, and his/her utility is $-\sqrt{5}$. However, if he/she lies and reports $(0, 1-\epsilon)$ instead, the outcome is $(0, 1-\epsilon)$ and his/her utility is $-(1+\epsilon) > -\sqrt{5}$ for any $0 \leq \epsilon \leq 1$. The mechanism does not satisfy the checks condition because, for example, the resulting outcome, $(1, 0)$, is worse than the status quo outcome, $(0, 0)$, for player G.

Taken together, the above examples illustrate a tension between strategy-proofness and Pareto efficiency in multidimensional policy space on the domain of single-peaked preferences. As already noted, this tension does not arise because the multidimensional moderate mechanism, in either of the above definitions, cannot satisfy both the properties. Previous work shows that strategy-proofness and efficiency cannot be simultaneously satisfied on the domain of preferences defined by Expression (1) (Border and Jordan, 1983; Kim and Roush, 1984). In effect, there is no mechanism that satisfies the checks, strategy-proofness, and efficiency conditions on the domain of Euclidean preferences, or on more general domains of single-peaked preferences that include the Euclidean preferences.

At the same time, it might be possible to find mechanisms that satisfy the checks and strategy-proofness conditions as well as weaker notions of efficiency. Consider the following weaker notion of efficiency, which we label minimal efficiency:

Minimal Efficiency: a mechanism $\mathbf{x}(p^C, p^G)$ is minimally efficient if and only if $p^C = p^G = p$ implies that $\mathbf{x}(p^C, p^G) = p$.

Minimal efficiency requires that, when all players have the same ideal policy, then the outcome will be that policy. Note that one can think of minimal efficiency as a weaker notion of efficiency because Pareto efficiency implies minimal efficiency.

Are there mechanisms that satisfy the strategy-proofness, minimal efficiency, and checks conditions in an n-dimensional policy space in which the players' preferences are represented by utilities of the form of Equation (1)? We get the following result:

Proposition 5 *The unique mechanism that satisfies the checks, strategy-proofness, and minimal efficiency conditions is* $\mathbf{m}(p^C, p^G)$.

Proof: See online appendix. ■

Proposition 5 shows that the unique mechanism that satisfies the strategy-proofness, minimal efficiency, and check conditions is the issue-by-issue moderate mechanism, $\mathbf{m}(p^C, p^G)$.[28] It suggests that when the equilibrium outcome of any checks and balances institution is not the moderate outcome, $\mathbf{m}(p^C, p^G)$, then the institution in question is either manipulable or not minimally efficient (or both).

Note that, in a unidimensional policy space, the set of mechanisms that are strategy-proof and Pareto efficient and the set of mechanisms that are strategy-proof and minimally efficient are equivalent.[29] Therefore, one can state a stronger version of Proposition 1: the unique mechanism that satisfies the checks, strategy-proofness, and minimal efficiency conditions in unidimensional policy space is the moderate mechanism.[30] Proposition 5 then is a generalization of this result to the multidimensional case.

Recall that in the unidimensional analysis, we replaced efficiency with responsiveness to show that the unidimensional moderate mechanism uniquely satisfies the checks, strategy-proofness, and responsiveness conditions. In the n-dimensional case, we showed above, the issue-by-issue moderate mechanism $\mathbf{m}(p^C, p^G)$ satisfies the strategy-proofness and checks conditions. The mechanism is also responsive, because given two profiles of ideal policies, (p^C, p^G) and (q^C, q^G), if $p^i_j > q^i_j$ for all $i \in \{C, G\}$ and all $j = 1, \ldots, n$, then $\mathbf{m}(p^C, p^G) > \mathbf{m}(q^C, q^G)$. Thus we have the following result:

Proposition 6 *The mechanism* $\mathbf{m}(p^C, p^G)$ *is strategy-proof and responsive and satisfies the checks condition.*

Proof: In text. ∎

Taken together, Propositions 5 and 6 suggest that the unique mechanism that satisfies the checks, strategy-proofness, minimal efficiency, and responsiveness conditions in the n-dimensional policy space is the issue-by-issue moderate mechanism, $\mathbf{m}(p^C, p^G)$. These results indicate that the welfare

[28] Note that we did not restrict this characterization to issue-by-issue mechanisms. In our characterization, we consider all multidimensional mechanisms, including those in which the outcome in a certain policy dimension may depend on the players' preferences on other dimensions.

[29] We prove this result in the context of the proof of Proposition 5.

[30] It is a stronger result because we would impose a weaker efficiency condition on the mechanism design characterization, while the Pareto efficiency condition is satisfied without requiring it.

properties of the moderate mechanism we identified in the unidimensional analysis are not driven by the dimensionality assumption.

Before concluding this section, we show that there is no tension between the checks and Pareto efficiency conditions on preference domains such as the one given by Expression (1).[31] Let T denote the set of policy outcomes that give every player a (weakly) higher payoff than the status quo outcome[32] and let E denote the set of efficient policy outcomes. Suppose that, contrary to the claim that there is no tension between the checks and efficiency conditions, and the set of mechanisms that satisfy both the checks and the efficiency conditions is empty, i.e., $T \cap E = \emptyset$. Then for any $y \in T$, we can find $z \in E$, such that $d(y, p^i) \geq d(z, p^i)$ for all $i \in \{C, G\}$ since the preference given by Expression (1) is continuous on $\mathbf{R}^n$. Because $T \cap E = \Phi$, $z \notin T$, and because $y \in T$, $d(y, p^i) \leq d(0, p^i)$ for all $i \in \{C, G\}$. These results, combined with the fact that $d(y, p^i) \geq d(z, p^i)$ for all $i \in \{C, G\}$, imply that $d(z, p^i) \leq d(0, p^i)$ for all $i \in \{C, G\}$, which in turn implies that $z \in T$, a contradiction. Therefore, $T \cap E \neq \emptyset$; mechanisms that satisfy both the checks and the Pareto efficiency conditions always exist.

5 Conclusions

In this paper, we use a mechanism design approach to investigate which, amongst all possible checks and balances institutions, are strategy-proof and efficient and strategy-proof and responsive. We show that checks and balances institutions in which the most preferred policy of the moderate player is the unique equilibrium outcome when players agree on the direction of policy change meet these criteria. Our approach facilitates taking any institutional setting where two or more players need to agree to effect policy change, formalizing that institution as a game, and then evaluating its equilibrium outcomes in accordance with the derived normative benchmark.

Our results can also be applied when evaluating institutional procedures in organizations that require the endorsements of all their members prior to making a decision. Coalition governance exemplifies such a situation (Laver and Shepsle, 1990). International organizations that emphasize the

[31] The preference domains on which the existing literature has shown that strategy-proofness and Pareto efficiency cannot be simultaneously satisfied.

[32] That is, $T \equiv \{y | d(y, p^i) \leq d(0, p^i) \forall i\}$, given an arbitrary profile of ideal policies (p^C, p^G). Note that T is not empty since $0 \in T$.

sovereignties of the member states and thus want to afford their members a capacity to block undesirable outcomes by employing unanimity rules are another example. For instance, the UN Security Council requires that all members approve specified courses of action before undertaking them; and the European Union Council of Minister requires unanimity on policy proposals that touch on vital interests of the member states, such as asylum, immigration, and foreign and security affairs.

Our analysis also has implications for a normative literature on the importance of moderation. Numerous political thinkers have viewed moderation as an important standard by which to judge and evaluate forms of governments. In *Politics*, Aristotle proposes moderation as an encompassing goal of political organization. In a similar vein, the principle of moderation animates Montesquieu's philosophy on good government in *Spirit of Laws*. Our paper contributes to this normative literature in showing that, within checks and balances institutions, the moderate's preferred policy is the unique outcome that satisfies a bundle of desirable properties such as strategy-proofness, efficiency, and responsiveness.

Numerous scholars have emphasized the importance of analyzing the designs of political institutions for the purpose of understanding them (Goodin, 1996). Although we focused specifically on the design of checks and balances institutions, students of political institutions can apply the mechanism approach we adopted here more generally, by first setting forth and justifying selected criteria and then determining which of all possible arrangements best meet them.[33] Because mechanism design facilitates a comprehensive evaluation of all possible institutional arrangements, given a set of criteria, it can foster the development of normative theories of political institutions.

References

Acemoglu, D., J. A. Robinson, and R. Torvik. 2013. "Why Do Voters Dismantle Checks and Balances?" *Review of Economic Studies* 80(3): 845–875.

Adler, M. 1998. "Rights Against Rules: The Moral Structure of American Constitutional Law." *Michigan Law Review* 97(1): 1–173.

Ashworth, S. and E. Bueno de Mesquita. 2007. " Monotone Comparative Statics for Models of Politics. " *American Journal of Political Science* 50(1): 214–231.

Banks, J. 1990. "Equilibrium Behavior in Crisis Bargaining Games." *American Journal of Political Science* 34(3): 599–614.

[33] Some mechanism design applications in political science are Banks (1990) and Baron (2000).

Barak, A. 2006. *The Judge in a Democracy.* Princeton, NJ: Princeton University Press.

Barbera, S. and M. Jackson. 1994. "A Characterization of Strategy-Proof Social Choice Functions for Economies with Pure Public Goods." *Social Choice and Welfare* 11(3): 241–252.

Baron, D. 2000. "Legislative Organization with Information Committees." *American Journal of Political Science* 44(3): 485–505.

Border, K. C. and J. S. Jordan. 1983. "Straightforward Elections, Unanimity and Phantom Voters." *Review of Economic Studies* 50(1): 153–170.

Brennan, G. and A. Hamlin. 2000. *Democratic Devises and Desires.* Cambridge, UK: Cambridge University Press.

Bueno de Mesquita, E. and M. Stephenson. 2007. "Regulatory Quality Under Imperfect Oversight." *American Political Science Review* 101(3): 605–620.

Buchanan, J. and G. Tullock. 1962. *The Calculus of the Consent. Logical Foundations of Constitutional Democracy.* Ann Arbor, MI: Michigan University Press.

Cameron, C. M. 2000. *Veto Bargaining: Presidents and the Politics of Negative Power.* Cambridge, MA: Cambridge University Press.

Cameron, C. M. and N. McCarty. 2004. "Models of Vetoes and Veto Bargaining." *Annual Review of Political Science* 7: 407–435.

Calvert, R. 1995. "Rational Actors, Equilibrium, and Social Institutions." In *Explaining Social Institutions*, Jack Knight and Itai Sened, (eds.), Ann Arbor, MI: University of Michigan Press, pp. 57–95.

Chung, K.-S. and J. C. Ely. 2006. "Ex-Post Incentive Compatible Mechanism." typescript.

Cox, G. W. and M. McCubbins. 2001. "The Institutional Determinants of Economic Policy Outcomes." In *Presidents, Parliaments and Policy*, Stephan Haggard and Mathew McCubbins, (eds.), Cambridge, UK: Cambridge University Press, pp. 21–65.

Cox, G. W. and M. D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives.* Cambridge, MA: Cambridge University Press.

Diermeier, D. and R. B. Myerson. 1999. "Bicameralism and Its Consequences for the Internal Organization of Legislatures." *The American Economic Review* 89(5): 1182–1196.

Dragu, T. and M. Polborn. 2013. "The Administrative Foundation of the Rule of Law." *Journal of Politics* 75(4): 1038–1050.

Ferejohn, J. and P. Pasquino. 2002. "Constitutional Courts as Deliberative Institutions." In *Constitutional Justice, East and West*, W. Sadurski, (ed.), The Hague: Kluwer Law International, pp. 21–36.

Ferejohn, J. and P. Pasquino. 2004. "Constitutional Adjudication: Lessons from Europe." *Texas Law Review* 82(7): 1671–1704.

Fox, J. and R. Van Weelden. 2010. "Partisanship and the Effectiveness of Oversight." *Journal of Public Economics* 94(9–10): 674–687.

Fox, J. and M. Stephenson. "The Welfare Effects of Minority-Protective Judicial Review." *Journal of Theoretical Politics.* In press.

Gailmard, S. and T. Hammond. 2011. "Intercameral Bargaining and Intracameral Organization in Legislatures." *Journal of Politics* 73(2): 535–546.

Fuller, L. L. 1969. *The Morality of Law.* 2nd revised edition. New Haven, CT: Yale University Press.

Goodin, R. E. (ed.) 1996. *The Theory of Institutional Design.* Cambridge, UK: Cambridge University Press.

Hayek, F. A. 1960. *The Constitution of Liberty.* Chicago, IL: The University of Chicago Press.

Hix, S., A. Noury, and G. Roland. 2006. *Democratic Politics in the European Parliament.* Cambridge, UK: Cambridge University Press.

Kim, K. H. and F. W. Roush. 1984. "Non-Manipulability in Two Dimensions." *Mathematical Social Sciences* 8(1): 29–43.

Keefer, P. and D. Stasavage. 2003. "The Limits of Delegation: Veto Players, Central Bank Independence and the Credibility of Monetary Policy." *American Political Science Review* 97(3): 407–423.
Kornhauser, L. 1992. "Modeling Collegial Courts. II. Legal Doctrine." *Journal of Law, Economics, & Organization* 8(3): 441–470.
Krehbiel, K. 1998. *Pivotal Politics: A Theory of U.S. Lawmaking*. Chicago, IL: University of Chicago Press.
Laver, M. and K. Shepsle. 1990. "Coalitions and Cabinet Government." *American Political Science Review* 84(3): 873–890.
Landes, W. and R. A. Posner. 1994. "The Economics of Anticipatory Adjudication." *Journal of Legal Studies* 23(2): 683–719.
Londregan, J. B. 2000. *Legislative Institutions and Ideology in Chile*. Cambridge, MA: Cambridge University Press.
Lupia, A. 1992. "Busy Voters, Agenda Control and the Power of Information." *American Political Science Review* 86(2): 390–403.
Manin, B. 1994. "Checks, Balances and Boundaries: The Separation of Powers in the Constitutional Debate of 1787." In *The Invention of the Modern Republic*, Biancamaria Fontana, (ed.), Cambridge, UK: Cambridge University Press, pp. 27–62.
Manin, B., A. Przeworski, and S. C. Stokes. 1999. "Elections and Representation." In *Democracy, Accountability, and Representation*, Adam Przeworski, Susan C. Stokes, and Bernard Manin, (eds.), Cambridge, UK: Cambridge University Press, pp. 29–54.
Matthews, S. A. 1989. "Veto Threats: Rhetoric in a Bargaining Game." *Quarterly Journal of Economics* 104(2): 347–369.
McCarty, N. 1997. "Presidential Reputation and the Veto." *Economics and Politics* 9(1): 1–27.
Melumad, N. and T. Shibano. 1991. "Communication in Settings with no Transfers." *Rand Journal of Economics* 22(2): 173–198.
Morgenstern, S. 2004. *Patterns of Legislative Politics: Roll Call Voting in the Latin America and the United States*. Cambridge, UK: Cambridge University Press.
Moulin, H. 1980. "On Strategy-Proofness and Single Peakedness." *Public Choice* 35(4): 437–455.
Myerson, R. B. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47(1): 61–73.
Myerson, R. B. 2008. "Mechanism Design." In *New Palgrave Dictionary of Economics* (2nd ed.), Steven N Durlauf and Lawrence E. Blume, (eds.), New York, NY: Palgrave Macmillan, pp. 533–542.
North, D. and B. Weingast 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *Journal of Economic History* 49(4): 803–832.
Ostrom, E. 2005. *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press.
Persson, T., G. Roland, and G. Tabellini. 1997. "Separation of Powers and Political Accountability." *Quarterly Journal of Economics* 112(4): 1163–1202.
Poole, K. and H. Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York, NY: Oxford University Press.
Primo, D. M. 2002. "Rethinking Political Bargaining: Policymaking with a Single Proposer." *Journal of Law Economics and Organization* 18(2): 411–437.
Rogers, J. and G. Vanberg. 2002. "Judicial Advisory Opinions and Legislative Outcomes in Comparative Perspective." *American Journal of Political Science* 46(2): 379–397.
Romer, T. and H. Rosenthal. 1978. "Political Resource Allocation, Controlled Agendas, and the Status Quo." *Public Choice* 33(4): 27–43.
Riker, W. H. 1982. *Liberalism against Populism: A Confrontation between the Theory of Democracy and the Theory of Social Choice*. San Francisco, CA: W.H. Freeman.
Shepsle, K. A. 1979. "Institutional Arrangements and Equilibria in Multidimensional Voting Models." *American Journal of Political Science* 23(1): 27–59.

Thayer, J. B. 1893. "The Origin and Scope of the American Doctrine of Constitutional Law." *Harvard Law Review* 7(3): 129–156.

Tushnet, M. 1999. *Taking the Constitution Away from the Courts*. Princeton, NJ: Princeton University Press.

Tsebelis, G. 2002. *Veto Players: How Political Institutions Work*. Princeton, NJ: Princeton University Press.

Voeten, E. 2000. "Clashes in the Assembly." *International Organization* 54(2): 185–215.

Weingast, B. 1997. "The Political Foundations of Democracy and the Rule of Law." *American Political Science Review* 91(2): 245–263.

The appendix contains two sections. In Appendix A, we present the proofs of our main results. In Appendix B, we extend our framework to a situation in which there are $N > 2$ decision-makers who must agree on changing an existing policy.

Appendix A

Proof of Lemma 1. First, we show that a strategy-proof mechanism $x(c, g)$ is weakly increasing in c for all g. Let us assume that for some g, $x(c, g)$ is not weakly increasing; i.e., without loss of generality, let $c_2 > c_1$ and $x(c_1, g) > x(c_2, g)$. Given this assumption, we consider the following four cases, which exhaust all possibilities:

(i) $c_1 \geq x(c_1, g) > x(c_2, g)$. Then $c_2 > x(c_1, g) > x(c_2, g)$; the player C with type c_2 prefers $x(c_1, g)$ to $x(c_2, g)$ and will have incentives to misrepresent its type. Thus $x(c, g)$ is not strategy-proof.

(ii) $x(c_1, g) > x(c_2, g) \geq c_1$. Then by the single-peakedness of $U_C(\cdot)$, the player C with type c_1 prefers $x(c_2, g)$ to $x(c_1, g)$ and will have incentives to misrepresent its type. Thus $x(c, g)$ is not strategy-proof.

(iii) $c_2 \geq x(c_1, g) > c_1 > x(c_2, g)$. Then by the single-peakedness of $U_C(\cdot)$, the player C with type c_2 prefers $x(c_1, g)$ to $x(c_2, g)$ and will have incentives to misrepresent its type. Thus $x(c, g)$ is not strategy-proof.

(iv) $x(c_1, g) > c_2 > c_1 > x(c_2, g)$. Incentive compatibility at c_1 implies that $U_C(x(c_1, g), c_1) \geq U_C(x(c_2, g), c_1)$. Single-crossing property further implies that $U_C(x(c_1, g), c_2) > U_C(x(c_2, g), c_2)$. As a result, the player C with type c_2 will prefer $x(c_1, g)$ to $x(c_2, g)$; thus $x(c, g)$ is not strategy-proof.

Second, we show that for arbitrary g, if $x(c, g)$ strictly increasing in c on an open interval (c_1, c_2), then $x(c, g) = c$ on (c_1, c_2). Suppose not. Without loss of generality consider the case in which $x(c^*, g) < c^*$ for some $c^* \in (c_1, c_2)$. If $x(c, g)$ is continuous at c^*, there exists an $\epsilon > 0$ such that $x(c^*, g) < x(c^* + \epsilon, g) < c^*$ and thus $U_C(x(c^* + \epsilon, g), c^*) > U_C(x(c^*, g), c^*)$ so $x(\cdot, g)$ is not strategy-proof.

If $x(c, g)$ is not continuous at c^*, since $x(c, g)$ is strictly increasing in c on (c_1, c_2), then $x(c, g)$ can only have jump discontinuities and there are at most countable discontinuity points. If $\mathbf{lim}_{c \to c^{*+}} x(c, g) < c^*$, we can use the preceding argument to show a contradiction (i.e., $x(\cdot, g)$ is not strategy-proof).

If $\mathbf{lim}_{c\to c^*+}x(c,g)=c^*$, since U_C is continuous there exists $\epsilon>0$ such that $U_C(x(c^*+\epsilon,g),c^*)>U_C(x(c^*,g),c^*)$ and thus $x(\cdot,g)$ is not strategy-proof.

If $\mathbf{lim}_{c\to c^*+}x(c,g)>c^*$, let us denote $\mathbf{lim}_{c\to c^*+}x(c,g)\equiv c^*+\Delta$ where $\Delta>0$. Then for any $c\in(c^*,\mathbf{min}\{c^*+\Delta,c_2\})$, we have $x(c,g)>\mathbf{lim}_{c\to c^*+}x(c,g)=c^*+\Delta>c$. Since $x(c,g)$ is strictly increasing on $(c^*,\mathbf{min}\{c^*+\Delta,c_2\})$, there exists $c^{**}\in(c^*,\mathbf{min}\{c^*+\Delta,c_2\})$ such that $x(c,g)$ is continuous at c^{**}, and from the previous line we also have $x(c^{**},g)>c^{**}$. Then there exists $\epsilon>0$ such that $x(c^{**},g)>x(c^{**}-\epsilon,g)>c^{**}$, and therefore $U_C(x(c^{**}-\epsilon,g),c^{**})>U_C(x(c^{**},g),c^{**})$. As a result, $x(\cdot,g)$ is not strategy-proof.

A similar argument holds if $x(c^*,g)>c^*$ for some $c^*\in(c_1,c_2)$.

As a result, for any g, any mechanism $x(c,g)$ that is strategy-proof for player C is weakly increasing in c, and if $x(c,g)$ is strictly increasing in c on an open interval (c_1,c_2), $x(c,g)=c$ on (c_1,c_2).

A similar argument can be used to characterize mechanisms that are strategy-proof for G, given an arbitrary type of C. ∎

Before proceeding to prove our main result, we first prove the following lemmas.

Lemma 2 *Let $x(c,g)$ be a strategy-proof mechanism. Then for any g, if $x(c,g)=c$ on (c_1,c_2), then $x(c,g)$ is continuous at both c_1 and c_2. Similarly, for any c, if $x(c,g)=g$ on (g_1,g_2), then $x(c,g)$ is continuous at both g_1 and g_2.*

Proof: We prove the first half of the lemma, and the proof for the second half is completely analogous. Without loss of generality, we prove that $x(c,g)$ is continuous at c_1. The case of c_2 is analogous. Suppose to the contrary that $x(c,g)$ is discontinuous at c_1. Since $x(c,g)$ is (weakly) increasing in c, $\mathbf{lim}_{c\to c_1-}x(c,g)<\mathbf{lim}_{c\to c_1+}x(c,g)=c_1$. Since U_C is continuous there exists small enough $\epsilon>0$ such that $x(c_1-\epsilon,g)\leq\mathbf{lim}_{c\to c_1-}x(c,g)<c_1-\epsilon$ and $U_C(x(c_1+\epsilon,g),c_1-\epsilon)>U_C(x(c_1-\epsilon,g),c_1-\epsilon)$. That is, the player C with type $c_1-\epsilon$ can do better by reporting $c_1+\epsilon$ and thus a contradiction to $x(c,g)$ being strategy-proof. ∎

In other words, the only discontinuity points of a strategy-proof mechanism $x(c,g)$ are the points that connect two flat segments. The next lemma further characterizes the discontinuity points of $x(c,g)$:

Lemma 3 *Let $x(c,g)$ be a strategy-proof mechanism. Then for any g, if $\hat{c}$ is a discontinuity point of $x(c,g)$, then $\hat{c} - \mathbf{lim}_{c\to\hat{c}^-} x(c,g) > 0$ and $\mathbf{lim}_{c\to\hat{c}^+} x(c,g) - \hat{c} > 0$. Similarly, for any c, if $\hat{g}$ is a discontinuity point of $x(c,g)$, then $\hat{g} - \mathbf{lim}_{g\to\hat{g}^-} x(c,g) > 0$ and $\mathbf{lim}_{g\to\hat{g}^+} x(c,g) - \hat{g} > 0$.*

Proof: Again we prove the first half of the lemma. Let us show that both terms are positive. There are only three other possibilities: (i) $\hat{c} - \mathbf{lim}_{c\to\hat{c}^-} x(c,g) \leq 0$ and $\mathbf{lim}_{c\to\hat{c}^+} x(c,g) - \hat{c} \leq 0$, but this is impossible because $x(c,g)$ is increasing and discontinuous at $\hat{c}$; (ii) $\hat{c} - \mathbf{lim}_{c\to\hat{c}^-} x(c,g) \leq 0$ and $\mathbf{lim}_{c\to\hat{c}^+} x(c,g) - \hat{c} > 0$; and (iii) $\hat{c} - \mathbf{lim}_{c\to\hat{c}^-} x(c,g) > 0$ and $\mathbf{lim}_{c\to\hat{c}^+} x(c,g) - \hat{c} \leq 0$. We show a contradiction for case (ii), and the argument for case (iii) is analogous.

Suppose $\hat{c} - \mathbf{lim}_{c\to\hat{c}^-} x(c,g) \leq 0$ and $\mathbf{lim}_{c\to\hat{c}^+} x(c,g) - \hat{c} > 0$. Since $x(c,g)$ is increasing and is discontinuous at $\hat{c}$, we have $\mathbf{lim}_{c\to\hat{c}^+} x(c,g) > \mathbf{lim}_{c\to\hat{c}^-} x(c,g) \geq \hat{c}$. Since U_C is continuous, there exists $\epsilon > 0$ such that $U_C(x(\hat{c}-\epsilon, g), \hat{c}+\epsilon) > U_C(x(\hat{c}+\epsilon, g), \hat{c}+\epsilon)$, so the player C with type $\hat{c}+\epsilon$ will have an incentive to misreport his/her type as $\hat{c}-\epsilon$ and thus $x(c,g)$ is not strategy-proof. ■

Lemma 4 *Let $x(c,g)$ be a strategy-proof mechanism. Then for any g, if for some $\hat{c}$, $x(\hat{c},g) = a \neq \hat{c}$, then $x(c,g) = a$ for all $c \in (\mathbf{min}\{a,\hat{c}\}, \mathbf{max}\{a,\hat{c}\})$. Similarly, for any c, if for some $\hat{g}$, $x(c,\hat{g}) = a \neq \hat{g}$, then $x(c,g) = a$ for all $g \in (\mathbf{min}\{a,\hat{g}\}, \mathbf{max}\{a,\hat{g}\})$.*

Proof: We prove the first half of the lemma for the case $a < \hat{c}$. The proof for $a > \hat{c}$ and the proof for the second half are analogous. Suppose there exists $\tilde{c} \in (a,\hat{c})$ such that $x(\tilde{c},g) \neq a$, then since $x(c,g)$ is (weakly) increasing, $x(\tilde{c},g) < a$. Thus $x(\tilde{c},g) < x(\hat{c},g) = a < \tilde{c}$, and since $U_C(\cdot,\tilde{c})$ is single-peaked with peak $\tilde{c}$, we have $U_C(x(\hat{c},g),\tilde{c}) > U_C(x(\tilde{c},g),\tilde{c})$, a contradiction to $x(c,g)$ being strategy-proof. ■

Proof of Proposition 1. By the definition of $m(c,g)$, if $c, g \geq 0$, then $m(c,g) = \mathbf{min}\{c,g\}$.

First, it is easy to check that the mechanism $x(c,g) = \mathbf{min}\{c,g\}$ satisfies the checks condition and strategy-proofness condition.

Next note that ex-post efficiency is equivalent to stating that for any (c,g), $x(c,g) \in [\mathbf{min}\{c,g\}, \mathbf{max}\{c,g\}]$. Therefore, the mechanism $x(c,g) = \mathbf{min}\{c,g\}$ is ex-post efficient and if $c = g$, the unique ex-post efficient outcome is $x(c,g) = c = g = \mathbf{min}\{c,g\}$.

Now suppose that for $c, g \geq 0$, there exists another mechanism $\tilde{x}(c, g)$ that satisfies all three conditions, such that for some $\tilde{c} \neq \tilde{g}$, $\tilde{c}, \tilde{g} \geq 0$, $\tilde{x}(\tilde{c}, \tilde{g}) \neq \mathbf{min}\{\tilde{c}, \tilde{g}\}$, and without loss of generality let $\tilde{c} < \tilde{g}$. Ex-post efficiency implies that $\tilde{x}(\tilde{c}, \tilde{g}) > \tilde{c}$.

Checks condition for player C with ideal point 0 implies that $\tilde{x}(0, \tilde{g}) = 0$. Lemma 1 then implies that $\tilde{x}(c, \tilde{g})$ has at least one discontinuous point on $[0, \tilde{c}]$. Now if $\tilde{x}(c, \tilde{g})$ is discontinuous at $c = 0$, i.e., if $\mathbf{lim}_{c \to 0^+} \tilde{x}(c, \tilde{g}) > 0$, then there exists $\epsilon > 0$ such that $U_C(0, \epsilon) > U_C(\tilde{x}(\epsilon, \tilde{g}), \epsilon)$. But then player C with ideal point ϵ has an incentive to deviate and report type 0, a contradiction to $\tilde{x}(c, g)$ being strategy-proof. Therefore, $\tilde{x}(c, \tilde{g})$ is continuous at $c = 0$ and therefore $\tilde{x}(c, \tilde{g})$ has at least one discontinuous point on $(0, \tilde{c}]$.

Let $\hat{c} \in (0, \tilde{c}]$ be a discontinuous point of $\tilde{x}(c, \tilde{g})$. From Lemma 3, we have $\hat{c} - \mathbf{lim}_{c \to \hat{c}^-} \tilde{x}(c, \tilde{g}) > 0$.

Therefore, there exists $c_1 \in (0, \hat{c}]$ such that $\tilde{x}(c_1, \tilde{g}) < c_1$. But since $c_1 \leq \hat{c} \leq \tilde{c} < \tilde{g}$, $\tilde{x}(c_1, \tilde{g})$ is not ex-post efficient. We have a contradiction.

Therefore, the unique mechanism that satisfies checks, strategy-proofness, and efficiency is $x(c, g) = \mathbf{min}\{c, g\} = m(c, g)$ for $c, g \geq 0$. ∎

Proof of Proposition 2. By the definition of $m(c, g)$, if $c, g \geq 0$, then $m(c, g) = \mathbf{min}\{c, g\}$.

First, it is easy to check that the mechanism $x(c, g) = \mathbf{min}\{c, g\}$ satisfies checks, strategy-proofness, and responsiveness.

Next, we show that any mechanism that is strategy-proof and responsive must be that $x(c, g) \in \{c, g\}$ for all (c, g). Suppose not, then there exist $(\hat{c}, \hat{g})$ and $a \notin \{\hat{c}, \hat{g}\}$ such that $x(\hat{c}, \hat{g}) = a$. Then $\mathbf{min}\{a, \hat{c}\} \neq \mathbf{max}\{a, \hat{c}\}$ and $\mathbf{min}\{a, \hat{g}\} \neq \mathbf{max}\{a, \hat{g}\}$. Lemma 4 implies that $x(c, \hat{g}) = a$ for all $c \in [\mathbf{min}\{a, \hat{c}\}, \mathbf{max}\{a, \hat{c}\}]$, and $x(\hat{c}, g) = a$ for all $g \in [\mathbf{min}\{a, \hat{g}\}, \mathbf{max}\{a, \hat{g}\}]$. This further implies that $x(c, g) = a$ for all $c \in [\mathbf{min}\{a, \hat{c}\}, \mathbf{max}\{a, \hat{c}\}]$ and $g \in [\mathbf{min}\{a, \hat{g}\}, \mathbf{max}\{a, \hat{g}\}]$. Therefore, $x(\mathbf{min}\{a, \hat{c}\}, \mathbf{min}\{a, \hat{g}\}) = x(\mathbf{max}\{a, \hat{c}\}, \mathbf{max}\{a, \hat{g}\}) = a$ but $\mathbf{max}\{a, \hat{c}\} > \mathbf{min}\{a, \hat{c}\}$ and $\mathbf{max}\{a, \hat{g}\} > \mathbf{min}\{a, \hat{g}\}$, a contradiction to $x(c, g)$ being responsive.

Therefore, if $x(c, g)$ is strategy-proof and responsive, $x(c, g) \in \{c, g\}$ for all (c, g). We next prove that for $c, g \geq 0$, the unique mechanism that satisfies checks, strategy-proofness, and responsiveness is $x(c, g) = \mathbf{min}\{c, g\}$. Suppose there exists another strategy-proof and responsive mechanism $\tilde{x}(c, g)$. From what we show above, since $\tilde{x}(c, g)$ is strategy-proof and responsive, $\tilde{x}(c, g) \in \{c, g\}$. For any $c = g$, $\tilde{x}(c, g) \in \{c, g\}$ implies $\tilde{x}(c, g) = \mathbf{min}\{c, g\}$, so there exists $\tilde{c} \neq \tilde{g}$ such that $\tilde{x}(\tilde{c}, \tilde{g}) \neq \mathbf{min}\{\tilde{c}, \tilde{g}\}$. Without loss of generality let $\tilde{c} < \tilde{g}$, then $\tilde{x}(\tilde{c}, \tilde{g}) = \tilde{g}$. Since $\tilde{x}(c, g)$ satisfies the checks condition, $\tilde{x}(0, \tilde{g}) = 0$. By Lemmas 1 and 3, $\tilde{x}(0, \tilde{g}) = 0$ and $\tilde{x}(\tilde{c}, \tilde{g}) = \tilde{g} > \tilde{c}$ imply that there exists $c_1 \in (0, \tilde{c}]$ such that $\tilde{x}(c, \tilde{g})$ is discontinuous at c_1, and there exists $c_2 < c_1$ such that $\tilde{x}(c_2, \tilde{g}) < c_2$. Since $c_2 < c_1 \leq \tilde{c} < \tilde{g}$, $\tilde{x}(c_2, \tilde{g}) \notin \{c_2, \tilde{g}\}$, a contradiction.

Therefore, the unique mechanism that satisfies checks, strategy-proofness, and responsiveness is $x(c, g) = \mathbf{min}\{c, g\} = m(c, g)$ for $c, g \geq 0$. ∎

Proof of Proposition 1′. First, it is easy to check that the mechanism $x(c, g) = m(c, g)$ if $c, g \geq 0$ or $c, g \leq 0$, and $x(c, g) = 0$ if $c < 0 < g$ or $g < 0 < c$, satisfies both the checks and (ex-post) Pareto efficiency conditions. To show that it is also strategy-proof on the general domain $\mathbf{R} \times \mathbf{R}$, it suffices to show the following three types of inequalities: (1) for $c, g \geq 0$, $U_C(x(c, g), c) \geq U_C(x(\tilde{c}, g), c)$ for any $\tilde{c} < 0$, and $U_G(x(c, g), g) \geq U_G(x(c, \tilde{g}), g)$ for any $\tilde{g} < 0$. The former is true because U_C is single-peaked and $x(c, g) = \mathbf{min}\{c, g\} \in [0, c]$ since $c, g \geq 0$ while $x(\tilde{c}, g) = 0$ since $\tilde{c} < 0 \leq g$. The latter inequality can be shown analogously; (2) for $c, g \leq 0$, $U_C(x(c, g), c) \geq U_C(x(\tilde{c}, g), c)$ for any $\tilde{c} > 0$, and $U_G(x(c, g), g) \geq U_G(x(c, \tilde{g}), g)$ for any $\tilde{g} > 0$. The proof is the same as in (1) above; and (3) for $c < 0 < g$ or $g < 0 < c$, $U_C(x(c, g), c) \geq U_C(x(\tilde{c}, g), c)$ for any $\tilde{c}$, and $U_G(x(c, g), g) \geq U_G(x(c, \tilde{g}), g)$ for any $\tilde{g}$. To show this, we show that when $c < 0 < g$, $U_C(x(c, g), c) \geq U_C(x(\tilde{c}, g), c)$ for any $\tilde{c}$, since all other cases are analogous. This is true because for any $\tilde{c} \leq 0$, $x(\tilde{c}, g) = x(c, g) = 0$, so $U_C(x(c, g), c) \geq U_C(x(\tilde{c}, g), c)$ trivially. Also, for any $\tilde{c} > 0$, $x(\tilde{c}, g) = \mathbf{min}\{\tilde{c}, g\} > 0$, so $U_C(x(c, g), c) \geq U_C(x(\tilde{c}, g), c)$ since U_C is single-peaked, $c < 0$, and $x(c, g) = 0$.

We next show that the mechanism $x(c, g) = m(c, g)$ if $c, g \geq 0$ or $c, g \leq 0$, and $x(c, g) = 0$ if $c < 0 < g$ or $g < 0 < c$ is the unique mechanism that satisfies checks, strategy-proofness, and efficiency on $\mathbf{R} \times \mathbf{R}$. Since $x(c, g) = m(c, g)$ is the unique mechanism that satisfies the three conditions on $\mathbf{R}_+ \times \mathbf{R}_+$, any mechanism that satisfies the three conditions on $\mathbf{R} \times \mathbf{R}$ must be equal to $m(c, g)$ on $\mathbf{R}_+ \times \mathbf{R}_+$. Similarly, any mechanism that satisfies the three conditions on $\mathbf{R} \times \mathbf{R}$ must be equal to $m(c, g)$ on $\mathbf{R}_- \times \mathbf{R}_-$, and equal to 0 on $(\mathbf{R}_- \times \mathbf{R}_+) \cup (\mathbf{R}_+ \times \mathbf{R}_-)$. ∎

Proof of Proposition 5. First, it is easy to check that the above issue-by-issue moderate rule is indeed strategy-proof, minimally efficient, and satisfies checks condition.

For uniqueness, first note that according to Border and Jordan (1983), a mechanism $\mathbf{x}(p^C, p^G)$: $\mathbf{R}^{2n} \to \mathbf{R}^n$ is strategy-proof and minimally efficient[1] if and only if there are mechanisms $x_1, x_2, \dots, x_n$: $\mathbf{R}^2 \to \mathbf{R}$ which are strategy-proof and minimally efficient such that

$$\mathbf{x}(p^C, p^G) = (x_1(p_1^C, p_1^G), x_2(p_2^C, p_2^G), \dots, x_n(p_n^C, p_n^G)),$$

where p_j^i is player i's ideal point on issue dimension j for $i \in \{C, G\}$ and $j = 1, 2, \dots, n$.

For any j, $x_j(p_j^C, p_j^G)$ is strategy-proof and minimally efficient, these imply that $x_j(p_j^C, p_j^G)$ is Pareto efficient, i.e., $\mathbf{min}\{p_j^C, p_j^G\} \le x_j(p_j^C, p_j^G) \le \mathbf{max}\{p_j^C, p_j^G\}$.[2] To show this, suppose first to the contrary that there exists (p_j^C, p_j^G) such that $x_j(p_j^C, p_j^G) > \mathbf{max}\{p_j^C, p_j^G\}$, without loss of generality suppose $p_j^C \le p_j^G$ so that $\mathbf{max}\{p_j^C, p_j^G\} = p_j^G$. But then $p_j^C \le x_j(p_j^G, p_j^G) = p_j^G < x_j(p_j^C, p_j^G)$, where the equality is due to minimal efficiency. This is a violation of strategy-proofness. Therefore, $x_j(p_j^C, p_j^G) \le \mathbf{max}\{p_j^C, p_j^G\}$. Similarly, we can show that $x_j(p_j^C, p_j^G) \ge \mathbf{min}\{p_j^C, p_j^G\}$, therefore, $x_j(p_j^C, p_j^G)$ is Pareto efficient.

Therefore, $\mathbf{x}(p^C, p^G)$ is strategy-proof and minimally efficient if and only if there exist $x_1, x_2, \dots, x_n$: $\mathbf{R}^2 \to \mathbf{R}$ which are strategy-proof and Pareto efficient such that $\mathbf{x}(p^C, p^G) = (x_1(p_1^C, p_1^G), x_2(p_2^C, p_2^G), \dots, x_n(p_n^C, p_n^G))$.

Next, we show that if $\mathbf{x}(p^C, p^G)$ satisfies the checks condition in addition to strategy-proofness and minimal efficiency, then the above $x_j(p_j^C, p_j^G)$ must also satisfy checks at 0 for all j, i.e., $x_j(p_j^C, p_j^G) = 0$ if $p_j^i = 0$ for some $i \in \{C, G\}$. Since $x(p^C, p^G)$ satisfies the checks condition, $\mathbf{x}(p^C, p^G) = \mathbf{0}$ if $p^i = \mathbf{0}$ for some $i \in \{C, G\}$. Let $p^C = \mathbf{0}$, then for any p^G, $\mathbf{x}(\mathbf{0}, p^G) = (x_1(0, p_1^G), x_2(0, p_2^G), \dots, x_n(0, p_n^G)) = \mathbf{0}$, therefore, $x_j(0, p_j^G) = 0$ for any j and any p_j^G. Similarly, $x_j(p_j^C, 0) = 0$ for any j and any p_j^C.

Therefore, $\mathbf{x}(p^C, p^G)$ is strategy-proof, minimally efficient, and satisfies the checks condition if and only if there exist $x_1, x_2, \dots, x_n$: $\mathbf{R}^2 \to \mathbf{R}$ which are strategy-proof, Pareto efficient, and satisfy checks at 0, such that $\mathbf{x}(p^C, p^G) = (x_1(p_1^C, p_1^G), x_2(p_2^C, p_2^G), \dots, x_n(p_n^C, p_n^G))$.

---

[1] In Border and Jordan (1983), minimal efficiency is labeled as unanimity.

[2] Note that $x_j(p_j^C, p_j^C)$ being Pareto efficient for all j is a necessary but not sufficient condition for $\mathbf{x}(p^C, p^G)$ to be Pareto efficient.

Finally, we show that if $\mathbf{x}(p^C, p^G)$ is strategy-proof, minimally efficient, and satisfies the checks condition, $\mathbf{x}(p^C, p^G) = (\mathbf{min}\{p_1^C, p_1^G\}, \mathbf{min}\{p_2^C, p_2^G\}, \ldots, \mathbf{min}\{p_n^C, p_n^G\})$. Since we have shown that $\mathbf{x}(p^C, p^G)$ is strategy-proof, minimally efficient, and satisfies the checks condition if and only if there exist $x_1, x_2, \ldots, x_n$: $\mathbf{R}^2 \rightarrow \mathbf{R}$ which are strategy-proof, Pareto efficient, and satisfy checks at 0, such that $\mathbf{x}(p^C, p^G) = (x_1(p_1^C, p_1^G), x_2(p_2^C, p_2^G), \ldots, x_n(p_n^C, p_n^G))$, it suffices to show that $x_j(p_j^C, p_j^G) = \mathbf{min}\{p_j^C, p_j^G\}$ for all $j = 1, \ldots, n$.

Consider an arbitrary $j \in \{1, 2, \ldots, n\}$. If $p_j^C = p_j^G$, minimal efficiency implies $x_j(p_j^C, p_j^G) = \mathbf{min}\{p_j^C, p_j^G\}$ trivially.

Suppose that there exists a mechanism $\tilde{x}_j(p_j^C, p_j^G) \neq \mathbf{min}\{p_j^C, p_j^G\}$ that will make $\mathbf{x}(p^C, p^G)$ satisfy the checks, strategy-proofness, and minimal efficiency conditions. That is, $\tilde{x}_j(p_j^C, p_j^G)$ is strategy-proof, Pareto efficient, and satisfies checks at 0, and for some $(\tilde{p}_j^C, \tilde{p}_j^G)$ such that $\tilde{p}_j^i \geq 0$ for $i \in \{C, G\}$, and $\tilde{p}_j^C \neq \tilde{p}_j^G$, $\tilde{x}_j(\tilde{p}_j^C, \tilde{p}_j^G) \neq \mathbf{min}\{\tilde{p}_j^C, \tilde{p}_j^G\}$. Pareto efficiency then implies that $\tilde{x}_j(\tilde{p}_j^C, \tilde{p}_j^G) > \mathbf{min}\{\tilde{p}_j^C, \tilde{p}_j^G\}$. Without loss of generality let $\tilde{p}_j^C \leq \tilde{p}_j^G$, so $\tilde{p}_j^C = \mathbf{min}\{\tilde{p}_j^C, \tilde{p}_j^G\}$.

Checks at 0 implies that $\tilde{x}_j(0, \tilde{p}_j^G) = 0$.

Since $\tilde{x}_j(p_j^C, p_j^G)$ is strategy-proof, Lemmas 1–4 in our one-dimension analysis applies. Lemma 1 then implies that $\tilde{x}_j(p_j^C, \tilde{p}_j^G)$ has at least one discontinuous point on $p_j^C \in [0, \tilde{p}_C]$. Now if $\tilde{x}_j(p_j^C, \tilde{p}_j^G)$ is discontinuous at $p_j^C = 0$, i.e., if $\mathbf{lim}_{p_j^C \to 0^+} \tilde{x}_j(p_j^C, \tilde{p}_j^G) > 0$, then there exists $\epsilon > 0$ such that $(0 - \epsilon)^2 < (\tilde{x}_j(\epsilon, \tilde{p}_j^G) - \epsilon)^2$. This means the player C with ideal point ϵ has an incentive to misreport to type 0, a contradiction to $\tilde{x}_j(p_j^C, p_j^G)$ being strategy-proof. Therefore, $\tilde{x}_j(p_j^C, \tilde{p}_j^G)$ is continuous at $p_j^C = 0$, and therefore $\tilde{x}_j(p_j^C, \tilde{p}_j^G)$ has at least one discontinuous point on $(0, \tilde{p}_j^C]$.

Let $\hat{p}_j^C \in (0, \tilde{p}_j^C]$ be a discontinuous point of $\tilde{x}_j(p_j^C, \tilde{p}_j^G)$. From Lemma 3, we have $\hat{p}_j^C - \mathbf{lim}_{p_j^C \to \hat{p}_j^{C-}} \tilde{x}_j(p_j^C, \tilde{p}_j^G) > 0$.

Therefore, there exists $q \in (0, \hat{p}_j^C]$ such that $\tilde{x}_j(q, \tilde{p}_j^G) < q$. But since $q \leq \hat{p}_j^C \leq \tilde{p}_j^C \leq \tilde{p}_j^G$, $\tilde{x}_j(q, \tilde{p}_j^G) < q \leq \mathbf{min}\{q, \tilde{p}_j^G\}$ and therefore is not Pareto efficient. We have a contradiction.

Therefore, if $\mathbf{x}(p^C, p^G)$ is strategy-proof, minimally efficient, and satisfies the checks condition, then

$$\mathbf{x}(p^C, p^G) = \mathbf{m}(p^C, p^G) = (\mathbf{min}\{p_1^C, p_1^G\}, \mathbf{min}\{p_2^C, p_2^G\}, \ldots, \mathbf{min}\{p_n^C, p_n^G\}).$$

■

Appendix B

In this section, we extend our main results to a situation in which there are $N > 2$ decision-makers who must agree on changing an existing policy. Similar to the main analysis, each of these N players has veto power in the sense that each has a reservation payoff given by the status quo policy, $q = 0$.

Formally, for $i = 1, 2, \ldots, N$, we denote the ideal policy of player i by p_i. A mechanism $x(p_1, p_2, \ldots, p_N)$ specifies an outcome $x \in X$ as a function of all members' reported types. A mechanism $x(p_1, p_2, \ldots, p_N)$ is a checks and balances mechanism if $U_i(x(p_1, p_2, \ldots, p_N), p_i) \geq U_i(q, p_i)$ for all i. We restate the following properties for the case of $N > 2$ players:

Checks: a mechanism $x(c, g)$ satisfies the checks condition if and only if $U_G(x(p_i, p_{-i}), p_i) \geq U_G(0, p_i)$ for all i, p_i, and p_{-i}.

Strategy-proofness: a mechanism $x(p_1, p_2, \ldots, p_N)$ is strategy-proof if and only if $U_i(x(p_i, p_{-i}), p_i) \geq U_i(x(\tilde{p}_i, p_{-i}), p_i)$ for all i, p_i, $\tilde{p}_i$ and p_{-i}.

Pareto Efficiency: a mechanism $x(p_1, p_2, \ldots, p_N)$ is (ex-post) Pareto efficient if and only if its outcome is Pareto efficient. Formally, for any $(p_1, p_2, \ldots, p_N)$, there is not another outcome x' such that $U_i(x', p_i) \geq U_i(x(p_1, p_2, \ldots, p_N), p_i)$ for all i, and $U_j(x', p_j) > U_j(x(p_1, p_2, \ldots, p_N), p_j)$ for some j.

Responsiveness: a mechanism $x(p_1, p_2, \ldots, p_N)$ is (strictly) responsive to players' preferences if and only if for any $(p'_1, p'_2, \ldots, p'_N)$ and $(p_1, p_2, \ldots, p_N)$ such that $p'_i > p_i \geq 0$ for all i, $x(p'_1, p'_2, \ldots, p'_N) > x(p_1, p_2, \ldots, p_N)$, and for any $(p'_1, p'_2, \ldots, p'_N)$ and $(p_1, p_2, \ldots, p_N)$ such that $p'_i < p_i \leq 0$ for all i, $x(p'_1, p'_2, \ldots, p'_N) < x(p_1, p_2, \ldots, p_N)$.

We also generalize the definition for the moderate player's most preferred outcome for the case of $N > 2$ players, $m(p_1, p_2, \ldots, p_N)$:

Definition *For any $(p_1, p_2, \ldots, p_N)$ where all p_i's have the same signs, the outcome $m(p_1, p_2, \ldots, p_N)$ is defined as the most preferred policy of the player whose ideal policy is the closest to the status quo $q = 0$; i.e., $m(p_1, p_2, \ldots, p_N) = \mathbf{min}_{i=1}^{N} p_i$ if $p_i \geq 0$ for all i, and $m(p_1, p_2, \ldots, p_N) = \mathbf{max}_{i=1}^{N} p_i$ if $p_i \leq 0$ for all i.*

Before proving the results with multiple decision-makers, we generalize Lemmas 1–4 to the model with N players.

Lemma 5 *For any i and any p_{-i}, any mechanism $x(p_i, p_{-i})$ that is strategy-proof for player i is weakly increasing in p_i, and if $x(p_i, p_{-i})$ strictly increasing in p_i on an open interval $({p_i}^1, {p_i}^2)$, $x(p_i, p_{-i}) = p_i$ on $({p_i}^1, {p_i}^2)$.*

Lemma 6 *Let $x(p_1, p_2, \ldots, p_N)$ be a strategy-proof mechanism. Then for any i and any p_{-i}, if $x(p_i, p_{-i}) = p_i$ on $({p_i}^1, {p_i}^2)$, then $x(p_i, p_{-i})$ is continuous at both p_i^1 and p_i^2.*

Lemma 7 *Let $x(p_1, p_2, \ldots, p_N)$ be a strategy-proof mechanism. Then for any i and any p_{-i}, if $\hat{p}_i$ is a discontinuity point of $x(p_i, p_{-i})$, then $\hat{p}_i - \mathbf{lim}_{p_i \to \hat{p}_i^-} x(p_i, p_{-i}) > 0$ and $\mathbf{lim}_{p_i \to \hat{p}_i^+} x(p_i, p_{-i}) - \hat{p}_i > 0$.*

Lemma 8 *Let $x(p_1, p_2, \ldots, p_N)$ be a strategy-proof mechanism. Then for any i and any p_{-i}, if for some $\hat{p}_i$ $x(\hat{p}_i, p_{-i}) = a \neq \hat{p}_i$, then $x(p_i, p_{-i}) = a$ for all $p_i \in (\mathbf{min}\{a, \hat{p}_i\}, \mathbf{max}\{a, \hat{p}_i\})$.*

The proof of the above lemmas are exactly the same as that of Lemmas 1–4.

Just as in the main analysis, we state the propositions for the case where $p_i \in \mathbf{R}_+$ for all i. We have the following results:

Proposition 7 *The unique mechanism that satisfies the checks, strategy-proofness, and Pareto efficiency conditions on $\mathbf{R}_+^N$ is $x(p_1, p_2, \ldots, p_N) = m(p_1, p_2, \ldots, p_N)$.*

Proof: When $p_i \geq 0$ for all i, $m(p_1, p_2, \ldots, p_N) = \mathbf{min}_{i=1}^N p_i$.

In our model with single-peaked preferences, an outcome x is efficient if and only if $\mathbf{min}_{i=1}^N p_i \leq x \leq \mathbf{max}_{i=1}^N p_i$.

First, it is easy to check that the mechanism $x(p_1, p_2, \ldots, p_N) = \mathbf{min}_{i=1}^N p_i$ satisfies checks, strategy-proofness, and efficiency.

Now, we show that for $(p_1, p_2, \ldots, p_N)$ where $p_i \geq 0$ for all i, $x(p_1, p_2, \ldots, p_N) = \mathbf{min}_{i=1}^N p_i$ is the unique mechanism that satisfies all the three conditions.

Note that if $p_i = p_j$, for all $i \neq j$, the unique efficient outcome is $x(p_1, p_2, \ldots, p_N) = \mathbf{min}_{i=1}^N p_i$.

Now suppose that there exists another mechanism $\tilde{x}(p_1, p_2, \ldots, p_N)$ that satisfies all the three conditions. That is, for some $(\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_N)$ such that $\tilde{p}_i \geq 0$ for all i, and $\tilde{p}_i \neq \tilde{p}_j$ for some $i \neq j$, $\tilde{x}(\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_N) \neq \mathbf{min}_{i=1}^N \tilde{p}_i$. Efficiency implies that $\tilde{x}(\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_N) > \mathbf{min}_{i=1}^N \tilde{p}_i$. Let $i^* \in \{1, 2, \ldots, N\}$ be such that $\tilde{p}_{i^*} = \mathbf{min}_{i=1}^N \tilde{p}_i$.

Checks condition for player i^* with ideal point 0 implies that $\tilde{x}(0, \tilde{p}_{-i^*}) = 0$. Lemma 5 then implies that $\tilde{x}(p_{i^*}, \tilde{p}_{-i^*})$ has at least one discontinuous point on $[0, \tilde{p}_{i^*}]$. Now if $\tilde{x}(p_{i^*}, \tilde{p}_{-i^*})$ is discontinuous at $p_{i^*} = 0$, i.e., if $\mathbf{lim}_{p_{i^*} \to 0^+} \tilde{x}(p_{i^*}, \tilde{p}_{-i^*}) > 0$, then there exists $\epsilon > 0$ such that $U_{i^*}(0, \epsilon) > U_{i^*}(\tilde{x}(\epsilon, \tilde{p}_{-i^*}), \epsilon)$. But then the player i^* with ideal point ϵ has an incentive to deviate and report type 0, a contradiction to $\tilde{x}(p_1, p_2, \ldots, p_N)$ being strategy-proof. Therefore, $\tilde{x}(p_{i^*}, \tilde{p}_{-i^*})$ is continuous at $p_{i^*} = 0$ and therefore $\tilde{x}(p_{i^*}, \tilde{p}_{-i^*})$ has at least one discontinuous point on $(0, \tilde{p}_{i^*}]$.

Let $\hat{p}_{i^*} \in (0, \tilde{p}_{i^*}]$ be a discontinuous point of $\tilde{x}(p_{i^*}, \tilde{p}_{-i^*})$. From Lemma 7, we have $\hat{p}_{i^*} - \mathbf{lim}_{p_{i^*} \to \hat{p}_{i^*}^-} \tilde{x}(p_{i^*}, \tilde{p}_{-i^*}) > 0$.

Therefore, there exists $p^1_{i^*} \in (0, \hat{p}_{i^*}]$ such that $\tilde{x}(p^1_{i^*}, \tilde{p}_{-i^*}) < p^1_{i^*}$. But since $p^1_{i^*} \le \hat{p}_{i^*} \le \tilde{p}_{i^*} \le \tilde{p}_j$ for all $j \neq i^*$, $\tilde{x}(p^1_{i^*}, \tilde{p}_{-i^*}) < p^1_{i^*} \le \mathbf{min}\{p^1_{i^*}, \tilde{p}_{-i^*}\}$ and therefore is not efficient. We have a contradiction.

Therefore, the unique mechanism that satisfies checks, strategy-proofness, and Pareto efficiency is $x(p_1, p_2, \ldots, p_N) = \mathbf{min}_{i=1}^N p_i = m(p_1, p_2, \ldots, p_N)$.

■

Proposition 8 *The unique mechanism that satisfies the checks, strategy-proofness, and responsiveness conditions on* $\mathbf{R}_+^N$ *is* $x(p_1, p_2, \ldots, p_N) = m(p_1, p_2, \ldots, p_N)$.

Proof: When $p_i \ge 0$ for all i, $m(p_1, p_2, \ldots, p_N) = \mathbf{min}_{i=1}^N p_i$.

First, it is easy to check that the mechanism $x(p_1, p_2, \ldots, p_N) = \mathbf{min}_{i=1}^N p_i$ satisfies checks, strategy-proofness, and responsiveness.

Next, we show that any mechanism that is strategy-proof and responsive must be that $x(p_1, p_2, \ldots, p_N) \in \{p_1, p_2, \ldots, p_N\}$ for all $(p_1, p_2, \ldots, p_N)$. Suppose not, then there exist $(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_N)$ and $a \notin \{\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_N\}$ such that $x(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_N) = a$. Then $\mathbf{min}\{a, \hat{p}_i\} \neq \mathbf{max}\{a, \hat{p}_i\}$ for all i. Lemma 8 implies that $x(p_i, \hat{p}_{-i}) = a$ for all i and $p_i \in [\mathbf{min}\{a, \hat{p}_i\}, \mathbf{max}\{a, \hat{p}_i\}]$. This further implies that $x(p_1, p_2, \ldots, p_N) = a$ for all $p_i \in [\mathbf{min}\{a, \hat{p}_i\}, \mathbf{max}\{a, \hat{p}_i\}]$. In particular, $x(\mathbf{min}\{a, \hat{p}_1\}, \mathbf{min}\{a, \hat{p}_2\}, \ldots, \mathbf{min}\{a, \hat{p}_N\}) = x(\mathbf{max}\{a, \hat{p}_1\}, \mathbf{max}\{a, \hat{p}_2\}, \ldots, \mathbf{max}\{a, \hat{p}_N\}) = a$ but $\mathbf{max}\{a, \hat{p}_i\} > \mathbf{min}\{a, \hat{p}_i\}$ for all i, a contradiction to $x(p_1, p_2, \ldots, p_N)$ being responsive.

Therefore, if $x(p_1, p_2, \ldots, p_N)$ is strategy-proof and responsive, $x(p_1, p_2, \ldots, p_N) \in \{p_1, p_2, \ldots, p_N\}$ for all $(p_1, p_2, \ldots, p_N)$. We next prove that for $(p_1, p_2, \ldots, p_N)$ where $p_i \ge 0$ for all i, the unique mechanism that satisfies checks, strategy-proofness, and responsiveness is $x(p_1, p_2, \ldots, p_N) = \mathbf{min}_{i=1}^N p_i$. Suppose there exists another mechanism $\tilde{x}(p_1, p_2, \ldots, p_N)$ that satisfies all the three conditions. From what we show above, since $\tilde{x}(p_1, p_2, \ldots, p_N)$ is strategy-proof and responsive, $\tilde{x}(p_1, p_2, \ldots, p_N) \in \{p_1, p_2, \ldots, p_N\}$ for all $(p_1, p_2, \ldots, p_N)$. For any $(p_1, p_2, \ldots, p_N)$ such that $p_i = p_j$ for all $i \neq j$, $\tilde{x}(p_1, p_2, \ldots, p_N) \in \{p_1, p_2, \ldots, p_N\}$ implies $\tilde{x}(p_1, p_2, \ldots, p_N) = \mathbf{min}_{i=1}^N p_i$. Therefore, there exists $(\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_N)$ with $\tilde{p}_i \neq \tilde{p}_j$ for some $i \neq j$, such that $\tilde{x}(\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_N) \neq \mathbf{min}_{i=1}^N \tilde{p}_i$. Let i^* be such that $\tilde{p}_{i^*} = \mathbf{min}_{i=1}^N \tilde{p}_i$, then $\tilde{x}(\tilde{p}_{i^*}, \tilde{p}_{-i^*}) > \tilde{p}_{i^*}$. Since $\tilde{x}(p_1, p_2, \ldots, p_N)$ satisfies the checks condition for player i^*, $\tilde{x}(0, \tilde{p}_{-i^*}) = 0$. By Lemmas 5 and 7, $\tilde{x}(0, \tilde{p}_{-i^*}) = 0$ and $\tilde{x}(\tilde{p}_{i^*}, \tilde{p}_{-i^*}) > \tilde{p}_{i^*}$ imply that there exists $p^1_{i^*} \in (0, \tilde{p}_{i^*}]$ such that $\tilde{x}(p_{i^*}, \tilde{p}_{-i^*})$ is discontinuous at $p^1_{i^*}$, and there exists $p^2_{i^*} < p^1_{i^*}$ such that $\tilde{x}(p^2_{i^*}, \tilde{p}_{-i^*}) < p^2_{i^*}$. Since $p^2_{i^*} < p^1_{i^*} \leq \tilde{p}_{i^*} \leq \tilde{p}_j$ for all $j \neq i^*$, $\tilde{x}(p^2_{i^*}, \tilde{p}_{-i^*}) \notin \{p^2_{i^*}, \tilde{p}_{-i^*}\}$, a contradiction.

Therefore, the unique mechanism that satisfies checks, strategy-proofness, and responsiveness is $x(p_1, p_2, \ldots, p_N) = \mathbf{min}_{i=1}^N p_i = m(p_1, p_2, \ldots, p_N)$. ■

Similar to what we found in the application to constitutional review section, the mechanism that implements the preferred policy of the more moderate player can be obtained as the unique equilibrium outcome by means of a simple institutional arrangement. In this institution, $N-1$ players choose a legal limit $\ell_i \in \mathbf{R}_+$ in a fixed sequential order,[3] and the remaining player chooses a policy $x \in \mathbf{R}_+$. A policy x is legal if and only if x does not exceed the lowest of the $N-1$ legal limits. In other words, a policy is legal if and only if it is within all the legal bounds set by the previous $N-1$ players. The outcome is x if x is legal and the status quo, $q = 0$, otherwise. We have the following proposition:

Proposition 9 *The unique equilibrium outcome in the institution in which $N-1$ players define the legal limits $\ell_i \in \mathbf{R}_+$ (in a fixed sequential order), and then the remaining player chooses a policy $x \in \mathbf{R}_+$, is $m(p_1, p_2, \ldots, p_N)$.*

Proof: Without loss of generality re-order the players such that the player with ideal points p_i is the ith one to make a choice. That is, players 1 through $N-1$ each chooses a legal limit in the order $1, 2, \ldots, N-1$, and then player

---

[3] The identities of these $N-1$ players can be arbitrary.

N (observing all the legal limits chosen by the previous $N-1$ players) chooses a policy. We claim that the strategy profile $\ell_i(\ell_1, \ldots, \ell_{i-1}, p_i) = p_i$ for $i = 1, 2, \ldots, N-1$ and $x_N(\ell_1, \ldots, \ell_{N-1}, p_N) = \mathbf{min}\{\mathbf{min}_{i=1}^{N-1} \ell_i, p_N\}$ is a Bayesian Nash equilibrium strategy and it gives as the unique Bayesian Nash equilibrium outcome the ideal policy of the most moderate player. We prove this claim by backward induction.

In the last stage, the effective legal limit is $\mathbf{min}_{i=1}^{N-1} \ell_i$, i.e., if player N chooses a policy not exceeding $\mathbf{min}_{i=1}^{N-1} \ell_i$ the outcome is this chosen policy, otherwise the outcome is the status quo. If the effective legal limit chosen by the previous $N-1$ players is $\mathbf{min}_{i=1}^{N-1} \ell_i$, for any strategy of the previous $N-1$ players and any beliefs of player N, $x_N(\ell_1, \ldots, \ell_{N-1}, p_N) = \mathbf{min}\{\mathbf{min}_{i=1}^{N-1} \ell_i, p_N\}$ is the unique optimal strategy for player N.

In the second-to-last stage, if the legal limits chosen by the previous $N-2$ players are $\ell_1, \ldots, \ell_{N-2}$, for any strategy of the previous $N-2$ players and any beliefs of player $N-1$, truth-telling (i.e., $\ell_{N-1}(\ell_1, \ldots, \ell_{N-2}, p_{N-1}) = p_{N-1}$) is optimal for player $N-1$. This is the case since if player $N-1$ deviates to $\ell' < p_{N-1}$, he/she either does not change the outcome and thus receives the same payoff, or changes the outcome from somewhere in between ℓ' and p_{N-1} to ℓ'. Since players have single-peaked preferences, this change will make player $N-1$ worse off. A similar argument shows that the player has no incentive for deviating to $\ell' > p_{N-1}$ either. Iterating this argument back to the first stage proves our claim. ∎