# Strategy-Proofness in Elections with Multidimensional Signals*

Justice C. Harasha†

March 15, 2024

**Abstract**

From Moulin's classic 1980 result, we know that, under a single-peaked domain, the Gibbard-Satterthwaite theorem can be sufficiently relaxed such that voters truthfully report their best preferred alternative when the central authority elects the median-reported 'peak.' This well-studied result provides an initial framing for this paper. We consider a two period election setting, where a policy is fixed in the first period. Agents report their ideal point and a *strength of preference* parameter, which denotes how sharply their utility decreases in movement to either direction of their ideal point. From this construction, we employ a mechanism design setting without transfers to consider the set of social choice functions that can be implemented in Bayesian-Nash equilibrium when agents are reporting this additional parameter. Then, after establishing the set of implementable social choice functions in a setting that does not allow transfers, we consider the same two period election setting where agents report an ideal point and strength of preference parameter, now allowing for monetary transfers and proceeding to characterize the set of social choice functions that are implementable in dominant strategies when transfers are allowed. We conclude with a brief discussion of application, namely considering how our results prove that it is in the best interest of party leaders and policymakers to consider the strength of voter preferences in movements away from ideal points and seek to give those with higher *strength of preference* parameters stronger sways over party directions and policy decision-making.

JEL CLASSIFICATION: C72, C73, D72, D82.

KEYWORDS: strategy proof mechanisms, voting games, political salience

*I acknowledge funding from the Yale University Department of Economics, the Yale Institution for Social and Policy Studies, and invaluable advisement from Adam Meirowitz and Dirk Bergemann. All errors are, of course, my own.

†Yale University, justice.harasha@yale.edu.

# Contents

# 1 Introduction

## 1.1 Relation to Literature

# 2 The Classical Setting: Single-Dimensional Reporting

In the forthcoming section, we familiarize the reader with the classic, single-dimensional environment, where voters strictly report an ideal point, and nothing more. This environment is well-studied, and the set of strategy-proof voting rules are characterized under appropriate technical assumptions by Moulin (1980). We offer a brief reconstruction of this general model and a summary of the main results unique to this environment, ultimately in an effort to demonstrate the fashion in which our novel multidimensional reporting differs from this classical construction.

## 2.1 The Single-Dimensional Model

Consider an environment with $n$ agents and a set $A$ of alternatives, or potential policies. We index the set of agents as follows: $i \in N = \{1, 2, ..., N\}$. We can think of each agent $i$ as a voter and the set $N$ as the electorate. Each voter $i$ has a value or utility function, $u_i(\theta)$, where $\theta \in \Theta$ is some unknown state. The state $\theta$ has $n$ elements $(\theta_1, \theta_2, ...\theta_n)$ and each agent $i$ observes only $\theta_i$, known as agent $i$'s type. In this setting, we refer to $\theta_i$ as voter $i$'s *true ideal point.* Simply, it denotes a given voter's truthfully preferred policy implementation over the set of possible alternatives or policies $A$. We assume that an ideal point $\theta_i$ is some real number drawn from a compact interval $\Theta_i$, an interval that we normalize to $[0, 1]$, without loss of generality. This yields the following construction of the normalized state space:

$$\Theta = \prod_{i=1}^{n} \Theta_i = [0, 1]^n. \tag{1}$$

We assume that states $\theta \in \Theta$ are distributed according to a probability measure $\mu(\cdot) \in \Delta(\Theta)$ that it has a continuously differentiable and strictly positive density function $f$.

The designer introduces a mechanism to decide on the collective policy implementation. We define a mechanism as the pair $(M, \psi)$, where $M$ is a collection of measurable message spaces $M_i$ and $\psi$ is an outcome function given by the mapping $\psi : M \to [0, 1]$, which assigns a policy choice for each message profile $m \in M$. In this setting, we call $M$ the set of *announced ideal points* and $\psi$ a *voting scheme*. Together, the mechanism $(M, \psi)$ induces a game of incomplete information where voters attempt to maximize their expected utilities.

This completes our reconstruction of the classical setting, which we call the single-dimensional model. We note that the single-dimensional aspect of the classical model arises from the fact that voters announce a single number, their announced ideal point, and no additional information is shared with the designer. In the next section, we revisit the main results from Moulin (1980) that characterize the set of incentive compatible voting schemes in this environment, and offer a strong basis upon which we build our multi-dimensional construction.

## 2.2 Single Peakedness and Strategy-Proof Voting Schemes

The most exhaustive characterization of this single-dimensional setting is done by Moulin (1980), who obtains the set of strategy-proof voting schemes in the classical environment, under appropriate restrictions. We offer this result, but first orient the reader with the set of restrictions on voter preferences and with formal notions of strategy-proof mechanisms.

To begin, Moulin assumes that for each voter $i$, their associated set of possible preferences $U_i$ is the set $S$ of *single-peaked preferences*, which we define below:

**Definition 1** *A preference profile $u$ is single-peaked if and only if there exists an alternative $a \in A$, called the peak of $u$, such that for all $x, y \in \mathbb{R}$*

$$x \leq y < a \implies u(x) \leq u(y) < u(a)$$

*and*

$$a < x \leq y \implies u(a) > u(x) \geq u(y)$$

It is important to note that we identify the preference preording $u$ with any utility or value function associated with it, as we initially defined $u(\cdot)$ Section 2.1.

From this notion of single-peakedness, we can offer the main solution concept that we consider in the classical environment: strategy-proofness, which we offer in two distinct definitions:

**Definition 2** *A voting rule $\psi$ is strategy proof if for every agent $i$ with single-peaked preference profile $u_i \in S$ and associated peak $a_i$, we have*

$$\forall x_i \in \mathbb{R},\ \ \forall \bar{x}_{-i} \in \mathbb{R}^{(n-1)} \quad u_i\bigg(\psi(a_i, \bar{x}_{-i})\bigg) \geq u_i\bigg(\psi(x_i, \bar{x}_{-i})\bigg)$$

*where $\bar{x}_{-i} \in \mathbb{R}^{(n-1)}$ is the $(n-1)$-uple of peaks announced by all other agents.*

From here, we also define the notion of group-strategy-proofness:

**Definition 3** *A voting rule $\psi$ is group-strategy-proof if for every coaltion $S \subset \{1, ..., n\}$ for every preference profile $(u_i)_{i \in S} \in S^S$ with associated peaks $a_S = (a_i)_{i \in S}$ we have that:*

$$\forall x_{S^C} \in \mathbb{R}^{S^C} \ \ \exists x_S \in \mathbb{R}^S \ \ \textit{such that } \forall i \in S$$

$$u_i\bigg(\psi(x_S, x_{S^C})\bigg) > u_i\bigg(\psi(a_S, x_{S^C})\bigg)$$

Simply, strategy-proofness guarantees that no agent has an incentive to report a signal that differs from his truthful, underlying peak, or type.

These are the primary notions of strategy-proofness that we will use in our review of Moulin's result in the classical single-dimensional environment as well in our forthcoming considerations of voting schemes in the novel multi-dimensional environment. We offer Moulin's central result below, without proof, which completes our characterization of the single-dimensional environment, and allows us to proceed to the multi-dimensional consideration:

**Theorem 1** *(Moulin, 1980). The statements below are equivalent:*

*(i) The voting rule $\psi : \mathbb{R}^n \to \mathbb{R}$ is group-strategy-proof and the selected policy implementation is Pareto optimal.*

*(ii) There exist $(n-1)$ real numbers $\alpha_1, ..., \alpha_{n-1} \in \mathbb{R} \cup \{+\infty, -\infty\}$ such that*

$$\forall (x_1, ..., x_n) \in \mathbb{R}^n \quad \psi(x_1, ..., x_n) = m(x_1, ..., x_n, \alpha_1, ..., \alpha_{n-1})$$

*where we recall that $m(x_1, ..., x_n, \alpha_1, ..., \alpha_{n-1})$ refers to the median value.*

Thus, from Moulin, we have a complete characterization of group-strategy-proof voting rules in the single-peaked, classical environment. We will employ this result several times during our characterization of the novel multivalued setting throughout Section 3.

# 3 Introducing the Multidimensional Environment

Suppose, now, that agents can report a multi-dimensional signal. That is, instead of simply announcing their peak, or ideal point $x_i$, agents report an additional parameter, say $\alpha_i$, that captures the strength of their preference for their reported peak. We formalize this addition to the classical setting by considering it in the form of a mechanism design problem.

## 3.1 A mechanism design problem

Consider an electorate composed of $n = 2p + 1$ agents, or voters. Each agent has a private value $\theta_i \in [0, 1] \times [\gamma, 1] := \Theta_i$, which is composed of her true peak $a_i$ and true *salience parameter*, $\bar{\alpha}_i$ which captures the strength of a given agent's preference for their peak. That is, an agent's private value or true type is of the form $\theta_i := (a_i, \bar{\alpha}_i)$. Agents then report a corresponding message $m_i \in [0,1] \times [\gamma, 1] := M_i$. Specifically, each agent $i$ reports an ordered pair of the form $(x_i, \alpha_i)$, where $x_i$ is agent $i$'s reported ideal point or announced peak and $\alpha_i$ is the agent's reported salience parameter. The timing of the mechanism is similar to the classical environment:

1. The designer, or policymaker, commits to the voting rule, which is given by the mapping $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$. The voting rule, in this case, is a measurable function that takes inputs of the agents' announced peaks and reported salience parameters, while outputting the selected policy implementation.

2. Agents announce their peaks $(x_1, ..., x_n)$ and salience parameters $(\alpha_1, ..., \alpha_n)$.

3. The designer collects these reports and enacts the voting rule $\psi$.

4. The outcome of the voting rule is selected as the implemented policy, $x^*_{imp}$.

From this multivalued construction, the policy implementation selected by the designer's voting rule can be a reflection not only of agent peaks or ideal points, but additionally of individualized preference intensities. In the next section, we provide a formal introduction to this salience parameter and the smaller subset of preference profiles that we initially restrict ourselves to.

## 3.2 On the salience parameter

We begin our analysis of the multivalued setting by restricting the our attention to a simple case of preference intensity and the salience parameter $\alpha_i$. Suppose that we assume agent preferences can be described by tent-shaped utility functions of the graphical form given below:

where an inverted absolute value function describes how an agent's announced peak corresponds to their highest utility and that their utility decreases symmetrically in movements

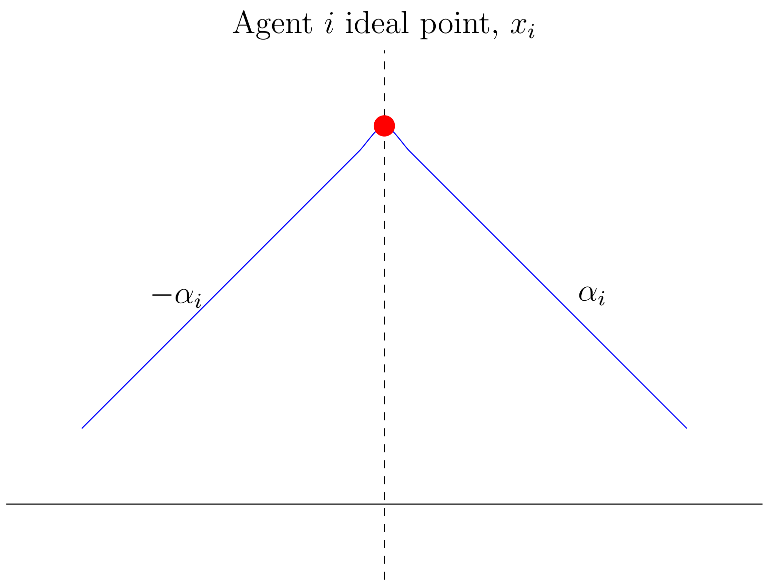


Figure 1: *An arbitrary voter's i simple utility function, centered at their ideal point $x_i$ (i.e. agent i's announced peak) with symmetric slopes given by $\alpha_i$.*

away from this ideal point in either direction. This is, of course, a highly simplified consideration of voter utility, but it nonetheless captures the intuition of strategy-proof voting rules in the multivalued environment. In Section 7, we consider a more general class of single-peaked preferences, namely quadratic utility functions. From here on, we call the general function given in Figure 1 *simple single-peaked (SSP)*. We offer a precise definition below:

**Definition 4** *A preference relation $u_i$ is called simple single-peaked (SSP) if it is of a symmetric tent shape centered at the agent's ideal point, of the exact functional form:*

$$u_i(x_i, \alpha_i) = -\Big|\alpha_i(x_i - x_0)\Big|.$$

*where $x_0$ is a constant which controls horizontal shifts of the tent-shaped utilities.*

From this initial grounding in the simple multivalued environment, we proceed by first considering a three agent direct mechanism setting. Later, we consider electorates of size $n \geq 3$ and indirect mechanisms.

### 3.3 A first-best benchmark

Suppose that $n = 3$ and that the type space coincides with the message space (i.e. $\Theta_i = M_i$). That is, we are restricting our attention to direct mechanisms, considering only the case where agents truthfully report their peak (ideal point) and salience parameter $(x_i, \alpha_i)$.

Thus, in a three-agent direct mechanism setting, we get the following three reports:

$$\Big\{(x_1, \alpha_1); (x_2, \alpha_2); (x_3, \alpha_3)\Big\}, \tag{2}$$

where each of the pairs correspond to the given agent's truthful peak and salience parameter. To aid in building some intuition in this modified multidimensional problem, we first characterize a notion of aggregate welfare maximization. That is, this portion of our analysis answers the very relaxed question: When the designer can observe the true ideal points and salience parameters of the agents, what policy should she select?

**Definition 5** *A given alternative $x^*_{FBWM}$ is called first-best welfare-maximizing (FBWM) if:*

$$x^*_{FBWM} \in arg\ max\Big\{\sum_{i=1}^{n} u_i(x_i, \alpha_i)\Big\}. \tag{3}$$

We proceed to characterize this notion in the three-agent setting. To do this, we offer a more precise construction of the environment. Without loss of generality, let us suppose that the follow ordering exists on the set of $\alpha_i$:

$$\alpha_1 \geq \alpha_2 \geq \alpha_3. \tag{4}$$

This yields two distinct possibilities:

$$\alpha_1 \geq \sum_{i \neq 1}^{n} \alpha_i \tag{5}$$

or

$$\alpha_1 \leq \sum_{i \neq 1}^{n} \alpha_i. \tag{6}$$

From (5) and (6), we arrive at a first result that begins our characterization of this environment:

**Lemma 1** *In a three-agent direct mechanism setting where (5) holds, the FBWM policy selection is* $\alpha_1$*'s truthfully reported peak,* $x_1$*, i.e.*

$$\psi\left[\left\{(x_1, \alpha_1); (x_2, \alpha_2); (x_3, \alpha_3)\right\}\right] = x_1 = x^*_{FBWM} \tag{7}$$

*is FBWM.*

**Proof.** We proceed by contradiction. That is, suppose that (5) holds, but that $x_1$ is not the first-best welfare-maximizing policy selection. This means that there exists some alternate policy selection, say $x'$, such that the following is true:

$$\sum_{i=1}^{n} u_i(x', \alpha_i) > \sum_{i=1} u_i(x_1, \alpha_i) \tag{8}$$

Explicitly, this yields:

$$u_1(x', \alpha_1) + u_2(x', \alpha_2) + u_3(x', \alpha_3) > u_1(x_1, \alpha_1) + u_2(x_1, \alpha_2) + u_3(x_1, \alpha_3) \tag{9}$$

By the simple single-peakedness of agent preferences, we know that the selection of $x_1$ yields the maximum value for voter 1's utility. We rewrite (9) in accordance with this fact:

$$u_1(x', \alpha_1) + u_2(x', \alpha_2) + u_3(x', \alpha_3) > \max\left\{u_1(x_1, \alpha_1)\right\} + u_2(x_1, \alpha_2) + u_3(x_1, \alpha_3) \tag{10}$$

We reorganize terms as follows:

$$\begin{aligned}&\left[u_1(x',\alpha_1)-\max\left\{u_1(x_1,\alpha_1)\right\}\right]\\&+\left[u_2(x',\alpha_2)-u_2(x_1,\alpha_2)\right]\\&+\left[u_3(x',\alpha_3)-u_3(x_1,\alpha_3)\right]>0\end{aligned}$$

We then proceed with term-by-term analysis. Clearly:

$$\left[u_1(x',\alpha_1)-\max\left\{u_1(x_1,\alpha_1)\right\}\right]<0 \tag{11}$$

since anything other than the maximum value is strictly less than the maximum value of a SSP utility function. Further, by construction, we know:

$$\left[u_2(x',\alpha_2)-u_2(x_1,\alpha_2)\right]>0, \tag{12}$$

since we view $x'$ as a an policy selection closer to the ideal points of voters 2 and 3. From this, we also get:

$$\left[u_3(x',\alpha_3)-u_3(x_1,\alpha_3)\right]>0 \tag{13}$$

We can normalize each term by the difference in ideal points by the fact that zero is on the right-hand side of the inequality, yielding:

$$\begin{aligned}&\frac{u_1(x',\alpha_1)-\max\left\{u_1(x_1,\alpha_1)\right\}}{x'-x_1}\\&+\frac{u_2(x',\alpha_2)-u_2(x_1,\alpha_2)}{x'-x_1}\\&+\frac{u_3(x',\alpha_3)-u_3(x_1,\alpha_3)}{x'-x_1}>0\end{aligned}$$

Each term is clearly equivalent to the slope between two points on each respective voter's utility function, which we know is simply their salience parameter $\alpha_i$. Then, using the sign derived from (11), (12), and (13), we have:

$$-\alpha_1 + \alpha_2 + \alpha_3 > 0 \tag{14}$$

$$\implies \alpha_2 + \alpha_3 > \alpha_1, \tag{15}$$

which is a clear contradiction of (5). Thus, $x_1$ is necessarily FBWM under the assumption in (5), completing our argument. ■

We proceed to consider the counterfactual in this initial setting, described by (6). Namely, we consider the question: What happens when the largest salience report is not all that much bigger than the rest of the reports? We get the following result:

**Lemma 2** *In a three-agent direct mechanism setting where (6) holds, the FBWM implementation is $\alpha_2$'s truthfully reported peak $x_2$ i.e.*

$$\psi\left[\left\{(x_1, \alpha_1); (x_2, \alpha_2); (x_3, \alpha_3)\right\}\right] = x_2 = x^*_{FBWM} \tag{16}$$

**Proof.** We, once more, proceed by contradiction. That is, suppose (6) holds, but that $x_2$ is not first-best welfare-maximizing. This means that there exists some alternate policy selection, say $x'$, such that the following is true:

$$\sum_{i=1}^{n} u_i(x', \alpha_i) > \sum_{i=1}^{n} u_i(x_2, \alpha_i) \tag{17}$$

$$\implies u_1(x', \alpha_1) + u_2(x', \alpha_2) + u_3(x', \alpha_3) > u_1(x_2, \alpha_1) + u_2(x_2, \alpha_2) + u_3(x_2, \alpha_3), \tag{18}$$

which we reorganize to yield:

$$\Big[u_1(x', \alpha_1) - u_1(x_2, \alpha_1)\Big\}\Big] + \Big[u_2(x', \alpha_2) - u_2(x_2, \alpha_2)\Big] + \Big[u_3(x', \alpha_3) - u_3(x_2, \alpha_3)\Big] > 0$$

From this arises two distinct cases:

(i) $x'$ is closer to the ideal point of voter 1 than $x_2$, and $x_2$ is closer to the ideal point of voter 3

(ii) $x'$ is closer to the ideal point of voter 3 than $x_2$, and $x_2$ is closer to the ideal point of voter 1

It is taken to be given that any ideal point other than $x_2$ is less desirable than $x_2$ for voter 2. From here, we consider the case described in (i). Since $x'$ is closer to the ideal point of voter 1 than $x_2$, we have:

$$\Big[u_1(x', \alpha_1) - u_1(x_2, \alpha_1)\Big\}\Big] > 0. \tag{19}$$

Further, since $x_2$ is the maximum value of voter 2's SSP utility, we have:

$$\Big[u_2(x', \alpha_2) - u_2(x_2, \alpha_2)\Big] < 0. \tag{20}$$

Finally, by (i) we know that $x_2$ is closer to the ideal point of voter 3, so we get:

$$\Big[u_3(x', \alpha_3) - u_3(x_2, \alpha_3)\Big] < 0. \tag{21}$$

After repeating a similar normalization procedure performed in the proof of Lemma 1 and the sign analysis performed in (19), (20), and (21), we get that:

$$\alpha_1 - \alpha_2 - \alpha_2 > 0 \tag{22}$$

$$\implies \alpha_1 > \alpha_2 + \alpha_3, \tag{23}$$

which is a clear contradiction of (6), and thus means that $x_2$ is FBWM in the case of (i).

We proceed to consider (ii). Since (ii) gives us the case where $x_2$ is closer to the ideal point of voter 2, we have:

$$\left[u_1(x', \alpha_1) - u_1(x_2, \alpha_1)\Big\}\right] < 0. \tag{24}$$

Since $x_2$ is still the maximum value of voter 2's SSP utility, (20) still holds. Finally, from (ii), we have that $x'$ is closer to the ideal point of voter 3 than $x_2$, so we have:

$$\left[u_3(x', \alpha_3) - u_3(x_2, \alpha_3)\right] > 0. \tag{25}$$

Again, by normalization and sign analysis, we get a convenient result in terms of $\alpha_i$ given below:

$$-\alpha_1 - \alpha_2 + \alpha_3 > 0 \tag{26}$$

$$\implies \alpha_3 > \alpha_1 + \alpha_2, \tag{27}$$

which is a clear contradiction since we have the following order on the $\alpha$ values: $\alpha_1 \geq \alpha_2 \geq \alpha_3$, and the inequality in (27) requires that the smallest salience parameter is greater than the sum of the two larger salience parameters. Therefore, under (ii), $x_2$ is also FBWM.

This thus completes our argument that $x_2$ is FBWM under (6). ■

We have thus characterized the three-agent environment when the designer observes the underlying types of each agent and can simply optimize the aggregation of their SSP utilities, in the two possible scenarios. In the forthcoming section, we proceed to consider an environment of incomplete information, considering the notion of dominant-strategy incentive compatibility, akin to Moulin's notion of strategy-proofness, and obtain a complete characterization of this setting with respect to that solution concept.

## 3.4 Strategy-proofness in three agent settings

In the previous section, we were concerned fundamentally with a question of unconstrained optimization: In the absence of any incentive compatibility constraints and endowed with the ability to observe each agent's true type, what should a policymaker or designer choose to select? We return to a more classical question, now, which includes the previously discussed notion of strategy-proofness. Suppose we are now interested in the implementation of a policy that makes it such that no agent has an incentive to report an ideal point and salience parameter that differs from her truthful, underlying type. In this way, we returnn to the notion of strategy-proofness and specify a notion of this in our multidimensional setting:

**Definition 6** *A voting scheme $\psi$ is called first-best strategy-proof (FBSP) if:*

$$\psi(x, \alpha) \in arg\ max\left\{\sum_{i=1}^{n} u_i(x_i, \alpha_i)\right\}. \tag{28}$$

*and*

$$\forall x_i \in \mathbb{R},\ \ \forall \bar{x}_{-i} \in \mathbb{R}^{(n-1)} \quad u_i\bigg(\psi(a_i, \bar{x}_{-i}), \alpha_i\bigg) \geq u_i\bigg(\psi(x_i, \bar{x}_{-i}), \alpha_i\bigg) \tag{29}$$

Simply, FBSP refers to the concept of the dominance of truth-telling as an optimal strategy. In the following result, we show that such an implementation is impossible in the three-agent multidimensional environment:

**Theorem 2** *In the three-agent environment, there exists no such FBSP voting scheme.*

**Proof.** We proceed by contradiction. Suppose that there exists such a FBSP voting scheme $\psi$, such that (28) and (29) are satisfied. From the results above, given in the case of the unconstrained problem (i.e. the case where the designer observes the true, underlying types of the voters), there are two possible solutions to (28). In the case where there is a single $\alpha_i$ report that satisfies: $\alpha_i \geq \sum_{j \neq i}^{n} \alpha_j$, where $\alpha_i \geq \alpha_j \geq \alpha_k$, we have shown that the welfare-maximizing policy selection is $x_i$. In the case case where $\alpha_i \geq \sum_{j \neq i}^{n} \alpha_j$, but $\alpha_i \leq \sum_{j \neq i}^{n} \alpha_j$, we have shown that the welfare-maximizing policy selection is $x_j$. Therefore, it is known in the three-agent setting that the first-best part of FBSP is only satisfied when $x_i$ or $x_j$ are the implementations, depending on the size of the largest $\alpha$ report. More specifically, this means that a FBSP voting scheme must select $x_i$ with some positive probability and $x_j$ with some positive probability such that the designer's objective function is maximized, as in (28).

This, however, means that there is an incentive for voters to misreport their $\alpha$ value, such that they report a salience parameter that is larger than the sum of the values of the salience parameters of their fellow agents, and thus ensure that their own ideal point is implemented. In this way, we can note that strategy-proofness, as given in (29), cannot always be satisfied when the designer's objective function is maximized, since a welfare-maximizing voting scheme creates an incentive for $\alpha$-deviation. When incentive for $\alpha$-deviation exists, truth-telling is no longer a dominant strategy, and thus strategy-proofness does not hold. ∎

# 4 Second-Best Policy Implementation in the Multidimensional Environment

As was shown above, a notion of first-best strategy-proofness does not exist in the multidimensional environment. Thus, in the forthcoming subsections, we will characterize a notion of *second-best implementation*, considering both pointwise incentive compatibility constraints (which is most representative of the classical strategy-proofness constraint) and introducing a notion of Bayesian implementation. This will complete our characterization of this baseline multidimensional environment, before we introduce multiple periods, additional agents, and the prospect of transfers, which all work to alleviate the stringent equilibrium conditions of the baseline environment.

## 4.1 Second-Best Strategy-Proofness

Our first consideration of second-best policy implementation is a notion of *second-best strategy-proofness*, which we obtain by constraining the policymaker's (principal) problem with pointwise incentive compatibility constraints. It is important to note that, while the pointwise nature of the constraints imply that agents do not form beliefs about each other's reports in this initial notion of second-best optimality, the principal or policymaker still forms beliefs about the aggregate expected payoff. This construction yields the following constrained optimization problem in the three-agent setting:

**Second-Best Maximization Problem 1** *Assume that each component of every agent's multi-dimensional report* $(x_i, \alpha_i)$ *are i.i.d. draws from the uniform distribution on the interval* $[0,1]$*, i.e.* $x_i, \alpha_i \overset{\text{iid}}{\sim} U[0,1]$, $\forall i \in N$. *Further, let* $x = (x_1, \ldots, x_n)$ *and* $\alpha = (\alpha_1, .., \alpha_n)$. *From this, we get the following policymaker's optimization problem:*

$$\sup_{\psi(\cdot)} \quad \int_{[0,1]} \overset{2n}{\cdots} \int_{[0,1]} \sum_{i=1}^{n} u_i \Big( x_i, \alpha_i, \psi(x,\alpha) \Big) dx_1 d\alpha_1 \ldots dx_n d\alpha_n \tag{30}$$

*subject to:*

$$u_i \Big[ x_i, \alpha_i, \psi \Big( (x_i, \alpha_i), (\bar{x}_{-i}, \bar{\alpha}_{-i}) \Big) \Big] \geq u_i \Big[ x_i, \alpha_i, \psi \Big( (x_i', \alpha_i'), (\bar{x}_{-i}, \bar{\alpha}_{-i}) \Big) \Big], \tag{31}$$

*where (31) is the strategy-proofness condition that holds for all* $x_i, \alpha_i, x_i', \alpha_i' \in \mathbb{R}$ *and for all* $\bar{x}_{-i}, \bar{\alpha}_{-i} \in \mathbb{R}^{n-1}$.

We begin our computation by evaluating and simplifying the objective function in (30). Upon integration and specifying the explicit functional form for each $i$, we arrive at:

(32)

# 5 Appendix