# Voting systems for Schelling games with discrete multiple choices (Draft)

## 1 Introduction

Social choice theory studies the ways in which the choices of many individuals between several possible outcomes can be translated into collective choices. In traditional elections, one typically imagines voters as having preferred choices; alternatively that they receive more or less utility depending on which candidate is chosen by the group. Then the motivation of participants is to vote in such a way that the resulting collective choice will provide them with as much utility as possible. Already this problem gives rise to a rich set of complexities; particularly, it has been seen that many natural properties that one would want electoral systems to have are incompatible [2], [11], [15], [14].

In this work we will consider a variant on this framework. Voters are still asked to provide ranked choices between a finite list of outcomes, but now they are indifferent to what outcome is selected. Instead, voters are financially rewarded or penalized based on the degree to which they agree with the rankings provided by other voters. Such systems have been proposed in the context of blockchain applications as a means of gaining access to off-chain information in the absence of a trusted authority, see [19], [21]. These systems might import onto a blockchain information such as which candidate won an election (for a prediction market platform) or whether it rained in Saskatchewan (for a crop insurance contract). The idea of this structure, which is based on the idea of a Schelling point (or focal point) [16], is typically that voters will select a list of choices that is "honest" as this is a distinguished choice and they expect other voters to also vote honestly. Hence, one is likely to be able to synthesize the votes of individuals incentivized in this way to produce outcomes that are correct most of the time.

In order for such Schelling game systems to produce useful results for these applications, one would want to structure them so that

- voters are incentivized as much as possible to *actually* provide an "honest" list of choices and
- the lists voters provide are converted into a collective outcome in such a way that if voters provide "honest" lists, the collective choice will also be "honest" (furthermore, ideally this conversion is robust in the event that some minority of voters provide malicious lists).

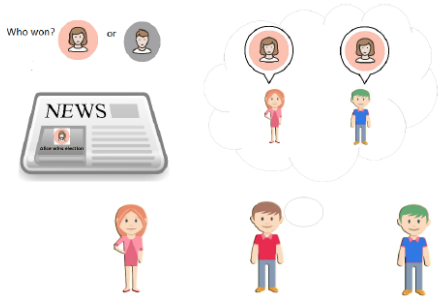


Figure 1: Each voter expect the other voters to submit the "honest" outcome as it is a distinguished choice. Hence as the voters are incentivized to be coherent, they have an incentive to vote honestly.

Hence, we want to choose a payoff structure for voters and a social choice function (i.e. a voting system) that interact in such a way as to achieve these goals.

Existing (blockchain-based) applications of this idea tend to use plurality voting. Often, the leading choice is so dominant (such as asking voters to report the winner of a well-known election), that even if voters are presented with many choices, plurality voting is not problematic. However, if voters are asked to choose between three or more choices that have a plausible chance of being selected by the group, plurality voting presents several issues that likely render it unacceptable for this setting:

- Voters have a harder time simulating how other voters will think - particularly one might expect different answers based on how many "levels deep" Alice goes in imaging how Bob thinks that Alice will think that Bob with think .... See Figure 2. This inhibits the formation of a Schelling point and leads to unpredictable outcomes.
- If the winning choice has support from substantially less than half of the voters, this lowers the threshold for an attacker tying to bribe voters/take control of the decision making process.

In this work we will propose a definition for what an "honest" vote can mean in this Schelling game framework. Then we will make observations on the choices of the payoff and voting systems in this model, in some cases providing heuristics on choices that we might expect to work well even in the presence of multiple possible outcomes. However, we will note that many of the subtleties of traditional social choice theory, specifically regarding the incompatibility of some natural, desirable properties, have analogs in this framework.

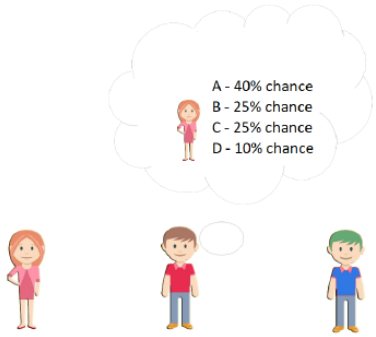


Figure 2: Bob imagines the probability of Alice's possible responses if she was polled on a question and, absent financial incentives, she gave her honest answer. As there is not a dominant choice, it becomes difficult to imagine how the other voters will think. Maybe they will all gravitate towards choice A as it would have plurality support in a poll. Maybe they gravitate to choice C thinking it is a reasonable compromise between A and B.

## 2 Related work

In a general sense, Schelling (or focal) points have a long tradition in social choice theory as a model for collaboration [17]. However, this "Schelling game" framework of having users vote on possible outcomes that have external consequences, using incentive structures based on whether the voters are coherent or incoherent so as to create a Schelling point around voting "honestly" seems to have only been proposed in recent years in the context of "oracles" for blockchain systems [19] [21].

Past work considering such Schelling games often focuses on situations where voters are presented with a binary choice (see for example [1]), avoiding both the classical problems of social choice theory in multi-candidate elections and the situation illustrated in Figure 2. To the degree that that systems such as [19] do present voters with non-binary choices, the "honest" choice is typically so clear (i.e. reporting the winner of a well-known election), that it is not problematic to choose the winner via a plurality voting system. In contrast, our current work is particularly motivated for applications to Kleros [13], a blockchain-based dispute resolution platform. Kleros, as a sort of "dispute resolution" oracle, operates under similar principles to other Schelling game systems, but it is designed to be able to handle cases where there is less likely to be unanimity around winning choices; in these cases, use of a plurality voting system may be less appropriate in the presence of non-binary outcomes.

In [10], a Schelling point based system that allows for non-binary answers, specifically that allows for real value answers, is considered, but this is done by binarizing the choices provided to voters. (Note that taking a binarization approach to multiple outcome decisions more generally has limitations, compare

to the discussion on discursive dilemmas in 17.1 of [4].) Work in standard social choice theory that more closely approaches this Schelling game framework, where voters are rewarded or penalized financially based on how they vote, is [18] that reconsiders the Gibbard-Satterthwaite Theorem in the context where one allows for divisible private goods (money) to be distributed among the voters in a way that is determined by the outcome of the vote.

# 3 Notation and basic notions

We use the following notation throughout: $A$ is a set of possible outcomes for a voting system. $L(A)$ is the set of possible orderings of the outcomes in $A$. $N$ is the number of voters.

It is natural to adapt certain ideas from standard social choice theory. Many properties of voting systems remain relevant, and we will consider them in our context, such as the Condorcet criterion, Condorcet winners, monotonicity, neutrality, and anonymity. See, for example, [4] for definitions of these properties. Furthermore, while it is possible that there are new, not yet considered, voting systems that are particularly adapted to the context of Schelling games, so far we have mostly studied using existing voting systems (combined with some payoff system). Particularly, in this work we have touched, in varying degrees of detail, on Plurality, Instant-Runoff, Alternative Smith (i.e. the system of checking if the Smith set is a unique outcome after each round of Instant-Runoff and if so returning that outcome as the winner), Ranked Pairs, and Dodgson; see [4] for descriptions of these systems.

For the purposes of this current draft, we generally avoid specifying how ties are dealt with; namely, we treat the systems we consider as if they are resolute. Our point of view is semi-justifiable in the context of Kleros, as voters are randomly drawn with probability proportional to their number of tokens [13]. As a result it is possible that some voters will have multiple votes; hence the resulting voting system is not anonymous (i.e. symmetric under permutations of the voters). In an extreme case (such as an appeal to the Kleros "general court") where the number of votes drawn is very large, one would expect each voter to have a number of votes closely aligned with the percentage of tokens she has staked. Then one could plausibly have situations where no subset of voters has exactly half of the weighted votes and hence no tie is even possible for any assignment of the participants' votes. Thus one avoids worrying about whether (exogenous) means of handling ties so that voting system are resolute results in failures of neutrality (compare to [5]). In future versions of this work, we will address in which situations we need to assume that voting systems are resolute.

# 4 Applicability of classical social choices theorems in this context

## 4.1 Arrow's Impossibility Theorem

**Theorem 1** (Arrow's Impossibility Theorem)**.** *Suppose* $\#A \geq 3$ *and* $F : L(A)^N \to L(A)$ *such that*

- *if* $r_i(a) > r_i(b)$ *for all* $i$ *then* $F(r_1, \dots r_N)(a) > F(r_1, \dots r_N)(b)$ *and*
- *if* $a$, $b$ *are in the same order according to* $r_i$ *and* $s_i$ *for all* $i$*, then* $a$, $b$ *have that order in* $F(r_1, \dots, r_N)$ *as in* $F(s_1, \dots s_N)$*.*

*Then* $F$ *is dictatorial.*

Note that Arrow's theorem only makes reference to the orderings provided by participants and the resulting ordering that the voting system produces. In particular, it does not depend on a user preferring the outcomes ranked according to their submitted ordering or otherwise consider the underlying incentive structure for the participants submitting those orderings. Hence, to the degree that any voting system we employ produces an ordering of possible outcomes in $L(A)$, Arrow's Theorem applies to our situation directly.

## 4.2 Gibbard-Satterthwaite Theorem

**Theorem 2** (Gibbard-Satterthwaite Theorem)**.** *Suppose* $\#A \geq 3$ *and* $f : L(A)^N \to A$ *is surjective such that if* $<_i, \dots <_N, <_i' \in L(A)$ *then*

$$a = f(<_1, \dots <_i \dots, <_N) >_i f(<_1, \dots <_i', \cdots <_N) = b.$$

*Then* $f$ *is a dictatorship.*

While this theorem, as stated in terms of the submitted orders $r_i$ of course still applies to our situation, its implications in terms of incentive compatibility do not directly apply. It is no longer the case that a user's goal is to submit the $r_i \in L(A)$ that results in the voting system choosing the $a \in A$ that is the highest value *according to* $r_i$. In Theorem 3, we will see a (preliminary) equivalent of Theorem 2 in our framework of Schelling game incentive structures.

# 5 Model

We consider the following model. A given user $\mathcal{USR}$ has some number of votes and she has a prior probability distribution $Vot(\mathcal{USR})$ for how all other voters will vote. (This is some given discrete probability distribution on the space $\Omega$ consisting of all possible assignments of the votes other than those of $\mathcal{USR}$.) The other voters' rulings are not necessarily identically distributed. For example, $\mathcal{USR}$ might know exactly how a given other voter will vote with probability one

while only having a probabilistic knowledge of others. Nor are the other voters' rulings necessarily independent of each other. Indeed $\mathcal{USR}$'s prior for their votes could include knowledge that voter $i$ will always vote exactly the same way as voter $j$ even if $\mathcal{USR}$ does not know either of their votes in advance. However, we assume that all of the other voters' rulings are independent of $\mathcal{USR}$'s vote. Then, whatever $Vot(\mathcal{USR})$ is and for any given vote by $\mathcal{USR}$, $\mathcal{USR}$ can estimate the probability of each outcome. Note that for applications to Kleros, there is the possibility of appeals of decisions. Here voters are incentivized based on whether they are coherent or incoherent with the answer in the last appeal round, rather than the majority in their own round [13]. So, when applicable, we also consider $\mathcal{USR}$'s estimation of the chance of appeal and the likelihood of results in eventual appeal rounds as being part of $Vot(\mathcal{USR})$.

Ideally, in situations like blockchain oracles where we think of there being a "true" outcome one would like voters to provide (a list of) votes based on what they genuinely believe that "true" outcome to be. Note that, in contrast to traditional social choice theory, voters in this framework no longer have underlying preferences for the outcomes; rather they are motivated by maximizing their economic return. Then, as is already the case in binary Schelling games, one cannot reward or punish voters with respect to whether they voted for the "true" outcome as the blockchain consensus protocol does not know which outcome this is, so instead the system rewards or punishes based on whether the user voted for the winning choice. In the case where there are multiple possible outcomes and voters provide a ranked list, a natural adaption of the Schelling game framework would be to ask them to list the outcomes in order of the outcome's chances of winning. One could hope that such lists will provide enough signal regarding the "truth" that a well-chosen social choice function will generally be able to use them to recover the "true" outcome.

**Definition 1.** *Given a prior probability distribution $Vot(\mathcal{USR})$ for how a user believes that other voters will vote which is independent from her own vote (also $Vot(\mathcal{USR})$ includes beliefs on the likelihood of appeals where applicable, though these are not necessarily independent of $\mathcal{USR}$'s vote), $\mathcal{USR}$'s vote $a_1 > a_2 > \ldots > a_M$ is said to be honest if*

$$p(a_1) \geq p(a_2) \geq \ldots \geq p(a_M),$$

*where $p(a_i)$ is the probability of $a_i$ winning based on $Vot(\mathcal{USR})$ and assuming that $\mathcal{USR}$ votes $a_1 > a_2 > \ldots > a_M$.*

In this work, we will consider this to be "honest" behavior and ask whether a given voting and reward/punishment system are incentive compatible. Note that there may be several honest votes possible to a user, to the degree that the user's vote changes the ordering of the most likely outcomes to win. Regardless, it is very natural to consider a vote which is not in order by probability as *dishonest*. The degree to which an incentive system encourages voters to submit such votes is a failure of incentive compatibility.

# 6 Candidate payoff systems

We begin this section with several versions of a criterion that a payoff system might satisfy. We will see in Proposition 1 that this idea is related to our definition of honesty in Definition 1.

**Definition 2.** *A payoff system that gives the payoff to $\mathcal{USR}_i$ according to*

$$G_i : L(A)^N \to \mathbb{R}$$

*is said to be reasonable if: given a priori $Vot(\mathcal{USR}_i)$ for $v_{-i}$, if $r_i$, $s_i \in L(A)$ are such that the cumulative probability of the first $j$ choices according to $r_i$ have greater than or equal to the probability of the first $j$ choices according to $s_i$ for all $j$, then $E[G_i(r_i, v_{-i})] \geq E[G_i(s_i, v_{-i})]$.*

When the relevant voter $\mathcal{USR}_i$ is clear from context, we will sometimes simply denote $G_i = G$.

**Definition 3.** *A reasonable payoff system that gives the payoff to $\mathcal{USR}_i$ according to*

$$G_i : L(A)^N \to \mathbb{R}$$

*is said to be strictly reasonable if: given a priori $Vot(\mathcal{USR}_i)$ for $v_{-i}$, if $r_i$, $s_i \in L(A)$ are such that the cumulative probability of the first $j$ choices according to $r_i$ have greater than or equal to the probability of the first $j$ choices according to $s_i$ for all $j$ and this cumulative probability is strictly greater for $r_i$ for at least one $j$, then $E[G_i(r_i, v_{-i})] > E[G_i(s_i, v_{-i})]$.*

**Lemma 1.** *Suppose $v_{-i}$ is some fixed set of choices for all votes other than those of $\mathcal{USR}$. Suppose $r_i$, $s_i \in L(A)$ are such that the winner $f(r_i, v_{-i})$ is ranked in the $k_1$st place of $r_i$ and the winner $f(s_i, v_{-i})$ is ranked in the $k_2$nd place of $s_i$. Then if the payoff system is reasonable, $k_1 \leq k_2 \Rightarrow G_i(r_i, v_{-i}) \geq G_i(s_i, v_{-i})$. Furthermore, if the payoff system is strictly reasonable, $k_1 < k_2 \Rightarrow G_i(r_i, v_{-i}) > G_i(s_i, v_{-i})$.*

*Proof.* Suppose that $\mathcal{USR}$'s prior is such that $v_{-i}$ has a 100% chance of occurring. Then the cumulative probability of the first $j$ choices for each of $r_i$ and $s_i$ is a step function in $j$ that passes from 0 to 1 at $k_1$ and $k_2$ respectively. Then the result follows directly, noting that as $v_{-i}$ has probability one of occurring, $E[G_i(r_i, v_{-i})] = G_i(r_i, v_{-i})$ and $E[G_i(s_i, v_{-i})] = G_i(s_i, v_{-i})$.

□

We consider another variant on this concept:

**Definition 4.** *A payoff system that gives the payoff to $\mathcal{USR}_i$ according to*

$$G_i : L(A)^N \to \mathbb{R}$$

*is said to be fatalistically reasonable if the following condition holds: Given a priori $Vot(\mathcal{USR}_i)$ for $v_{-i}$ and such that $Vot(\mathcal{USR}_i)$ is such that with probability*

1, *the voting system chooses the same result regardless of the vote of* $\mathcal{USR}$. *Then if* $r_i$, $s_i \in L(A)$ *are such that the cumulative probability of the first* $j$ *choices according to* $r_i$ *have greater than or equal to the probability of the first* $j$ *choices according to* $s_i$ *for all* $j$, *then* $E[G_i(r_i, v_{-i})] \geq E[G_i(s_i, v_{-i})]$.

Note that every reasonable incentive system is fatalistically reasonable. While this fatalistic framework may not be very interesting from the perspective of traditional social choice theory, in the context of Kleros, it is fairly natural. In Kleros, one can believe that unjust outcomes will be appealed so that final outcomes are relatively detached from the votes in a given voting round.

**Proposition 1.** *Suppose that we have a fatalistically reasonable incentive system. Suppose that the user's priors are such that with probability* 1, *the voting system chooses the same result regardless of the vote of* $\mathcal{USR}$. *Then* $\mathcal{USR}$ *maximizes her expected payoff with an honest vote.*

*Proof.* First note that, by our assumptions, the probability of each outcome winning does not depend on $\mathcal{USR}$'s vote. Hence we can list the outcomes by their probability of winning. In particular, this ordering $r_i$ is an honest vote. Moreover, the first $j$ choices of $r_i$ have cumulative probability at least equal to the cumulative probability of any $s_i$ as the choices in the first $j$ outcomes of $r_i$ correspond to those chosen by a greedy algorithm.

□

Suppose that all voters submit an ordering in $L(A)$ along with a deposit $D$. We now consider a possible (simplified) incentive system.

**Definition 5.** *(Toy incentive system)*

- *Determine the winner* $w \in A$ *via an underlying voting system.*
- *For each pair of the winner versus another choice in* $A$, *e.g.* $w$ *versus* $a$, $w$ *versus* $b$, *etc any voter who did not rank* $w$ *ahead of the other choice loses a deposit* $d = \frac{D}{\#A-1}$.
- *These lost deposits are burnt.*

**Proposition 2.** *The toy incentive system is reasonable.*

*Proof.* Suppose $r_i$ consists of $a_1 > a_2 > a_3 > ... > a_L$ Then

$$E(\text{payoff of } r_i) = \sum_k -(k-1) \cdot \text{prob}(a_k \text{ wins}).$$

Suppose $s_i$ consists of $b_1 > b_2 > b_3 > ... > b_L$ and that

$$\sum_{l=1}^{j} \text{prob}(a_l \text{ wins}) \geq \sum_{l=1}^{j} \text{prob}(b_l \text{ wins})$$

for all $j$.

$$\sum_k (k-1)\cdot \text{prob}(a_k \text{ wins}) = \sum_{j=2}^{L}\sum_{k=j}^{L} \text{prob}(a_k \text{ wins}) = \sum_{j=2}^{L}\left(1 - \sum_{k=1}^{j-1} \text{prob}(a_k \text{ wins})\right).$$

Hence

$$\sum_k -(k-1)\cdot \text{prob}(a_k \text{ wins}) \geq \sum_k -(k-1)\cdot \text{prob}(b_k \text{ wins})$$

$$\Leftrightarrow \sum_{j=2}^{L}\sum_{k=1}^{j-1} \text{prob}(a_k \text{ wins}) \geq \sum_{j=2}^{L}\sum_{k=1}^{j-1} \text{prob}(b_k \text{ wins})$$

(where we have canceled $\sum_{j=2}^{L} 1$ from each side). This holds by the assumption of reasonableness.

□

Note in Kleros [13], if one offered arbitration fee payouts to all voters, coherent or otherwise, and then burned lost deposits of PNK (the Kleros token used as deposits by arbitrators on a dispute), this would be equivalent to the toy payoff system (as the constant fee payouts do not alter incentives).

We now consider a more realistic incentive system.

**Definition 6** (Candidate payoff system)**.**

- *Determine the winner $w \in A$ via an underlying voting system.*
- *For each pair of the winner versus another choice in $A$, e.g. $w$ versus $a$, $w$ versus $b$, etc any voter who did not rank $w$ ahead of the other choice loses a deposit $d = \frac{D}{\#A-1}$.*
- *Voters are paid a reward for their coherent votes of*

  $$\frac{\#\text{ total incoherent votes across all pairs}}{\#\text{ total coherent votes across all pairs}} \#\text{ pairs in which } \mathcal{USR} \text{ coherent} \cdot d.$$

  *Similarly, arbitration fees can also be paid as*

  $$\frac{\text{Total arbitration fees}}{\#\text{ total coherent votes across all pairs}} \#\text{ pairs in which } \mathcal{USR} \text{ coherent.}$$

**Proposition 3.** *The candidate payoff system is fatalistically reasonable.*

*Proof.* Suppose that the user's priors are such that with probability 1, the voting system chooses the same result regardless of the vote of $\mathcal{USR}$. Denote by $K$ the total number of pairwise votes on which users other than $\mathcal{USR}$ are coherent with the ultimate winning choice. Similarly, denote by $K'$ the total number of deposits $d$ lost by users other than $\mathcal{USR}$. Denote by $F$ the arbitration fees

paid to be split between the voters of this round. Take votes $a_1 > \ldots a_L$ and $b_1 > \ldots > b_L$ such that

$$\sum_{l=1}^{j} \text{prob}(a_l \text{ wins}) \geq \sum_{l=1}^{j} \text{prob}(b_l \text{ wins})$$

for all $j$.

Note that, based on our assumptions, the ultimate winning choice can be considered to be fixed. Then for any given vote $v_{-i}$ of the other voters, $K$ and $K'$ are also fixed.

Suppose $\mathcal{USR}$ places the ultimate winning choice in the $i$th position. Then she is correct regarding $L-1-(i-1) = L-i$ pairs and incorrect regarding $i-1$ pairs. Hence, her payoff, accounting for lost deposits is:

$$\text{payoff}(i) = \left([K' + (i-1)]\, d + F\right) \frac{L-i}{K+L-i} - (i-1)d.$$

Note that for any lost deposit, $\mathcal{USR}$ can only recover a portion of it through her reward so

$$\text{payoff}(i) \geq \text{payoff}(i+1).$$

Then

$$E[\text{vote } a_1 > a_2 > ... > a_L] = \sum_{j=1}^{L} \text{payoff}(j)\text{prob}(a_j \text{ wins})$$

$$= \sum_{j=1}^{L-1} \text{payoff}(j)\text{prob}(a_j \text{ wins}) + \text{payoff}(L)\text{prob}(a_L \text{ wins})$$

$$= \sum_{j=1}^{L-1} \text{prob}(a_j \text{ wins})\text{payoff}(j) - \text{payoff}(L)\left(1 - \text{prob}(a_L \text{ wins})\right) + \text{payoff}(L)$$

$$= \text{payoff}(L) + \sum_{j=1}^{L-1} \text{prob}(a_j \text{ wins})\text{payoff}(j) - \text{payoff}(L) \sum_{j=1}^{L-1} \text{prob}(a_j \text{ wins})$$

$$= \text{payoff}(L) + \sum_{j=1}^{L-1} \text{prob}(a_j \text{ wins})\left(\text{payoff}(j) - \text{payoff}(L)\right)$$

$$= \text{payoff}(L) + \sum_{j=1}^{L-1} \text{prob}(a_j \text{ wins}) \sum_{i=j}^{L-1} [\text{payoff}(i) - \text{payoff}(i+1)]$$

$$= \text{payoff}(L) + \sum_{j=1}^{L-1} \sum_{i=j}^{L-1} [\text{payoff}(i) - \text{payoff}(i+1)]\text{prob}(a_j \text{ wins})$$

$$= \text{payoff}(L) + \sum_{i=1}^{L-1} \sum_{j=1}^{i} [\text{payoff}(i) - \text{payoff}(i+1)] \text{prob}(a_j \text{ wins})$$

$$= \text{payoff}(L) + \sum_{i=1}^{L-1} \left( [\text{payoff}(i) - \text{payoff}(i+1)] \sum_{j=1}^{i} \text{prob}(a_j \text{ wins}) \right).$$

Then, by our assumptions on the votes, this is greater than or equal to

$$\text{payoff}(L) + \sum_{i=1}^{L-1} \left( [\text{payoff}(i) - \text{payoff}(i+1)] \sum_{j=1}^{i} \text{prob}(b_j \text{ wins}) \right) = E[\text{vote } b_1 > b_2 > ... > b_L].$$

□

## 6.1 Effect of multiple choices on costs of attacks

We consider the impact of this payoff system on potential attacks. For example, for a bribe attack, we imagine that an attacker must offer a bribe larger than

$$\text{moral cost of accepting bribe} + d + \text{opportunity costs of voting honestly.}$$

Note that in Kleros the rewards, both the redistribution of PNK and the payment of ETH, are of the form

$$\frac{\text{Amount}}{\#\text{ total coherent votes across all pairs}} \#\text{ pairs in which } \mathcal{USR} \text{ coherent}$$

$$= \frac{\text{Amount}}{\text{average \#coherent pairs per voter } \cdot \#\text{voters}} \#\text{ pairs in which } \mathcal{USR} \text{ coherent}$$

If a voter accepts a bribe that cause them to be coherent on one less pair, as is the case if they rank a given attacker choice ahead of the honest choice but otherwise rank honestly, she loses

$$\frac{\text{Amount}}{\text{average \#coherent pairs per voter } \cdot \#\text{voters}}$$

of these rewards as well as a deposit $d$.

Roughly/heuristically (and depending on how moral costs of accepting the bribe are affected) the penalty for accepting the bribe is $\frac{1}{\#A-1}$ what it would be if there were only two choices.

## 6.2 Weighting contentious decisions more

One way to reduce the dilution of the cost on accepting bribes in terms of lost deposits/rewards and to increase the cost of an attack is to weight the pairs on which $\mathcal{USR}$ is coherent or not so that pairs that are very narrowly decided are weighted more. (Again, for our purposes, whether $\mathcal{USR}$ is coherent or not on a pairs means whether $\mathcal{USR}$ placed the eventual winner ahead of the other possible outcome rather than whether $\mathcal{USR}$ agreed the raw pair-wise preferences between those two outcomes).

This is in the spirit that *narrowly* failed attacks should be particularly expensive, which is a common goal in the design of blockchain-based platforms [6]. If an attacker is attempting to commit a sufficient number of bribes so that a would-be Condorcet winner no longer wins, this requires at least one pair passing from the honest winner winning to not; moreover, in the Ranked Pairs system, if two results $a$ and $b$ are both possible based on the way some small number of attacker votes are cast, that means that the pair between $a$ and $b$ must be ranked low enough to either be reversed or not be included in one of the two outcomes. Hence, as a borderline attack increases its number of votes controlled, it passes through a stage where it has just enough votes so that a pair involving the (honest) winner is decidedly narrowly. By weighting these pairs heavily, by narrowly failing, an attacker loses a larger percentage of her deposits.

Then, taking $\alpha > 0$, we define weights:

$$w(i) = \frac{\left(\frac{1}{|\text{margin of } a_i \text{ against winner}|+1}\right)^{\alpha}}{\sum_{a_j \text{ not winner}} \left(\frac{1}{|\text{margin of } a_j \text{ against winner}|+1}\right)^{\alpha}}.$$

Now we have voters lose

$$D \sum_{a_j \text{ not winner}} \mathbf{1}_{\text{voter voted } a_j \text{ ahead of winner}} \cdot w(j)$$

from their deposit $D$ and receive redistributions of the form

$$\frac{\text{Amount}}{\sum_{\text{voter } k} \sum_{a_j \text{ not winner}} \mathbf{1}_{\text{voter voted } a_j \text{ ahead of winner}} \cdot w(j)} \sum_{a_j \text{ not winner}} \mathbf{1}_{\mathcal{USR} \text{ voted } a_j \text{ ahead of winner}} \cdot w(j)$$

Here the $\alpha$ parameter allows one to tune how much narrowly decided pairs are weighted compared to less narrowly decided pairs. For example, if $\alpha = 0$, all weights are the same and we recover the previous incentive system 6; if $\alpha \to \infty$ then all the weight is concentrated on the narrowest pair (which is probably undesirable because, even though our comments above argue that a narrowly failed attack will lose some narrow pair, this pair may not be a priori the narrowest, so placing all the weight on the narrowest pair may actually make the attack cheaper). The "plus 1"s here are to allow $w(i)$ to be defined even if some pairs are tied.

**Example 1.** *Weighting the pairs causes our system to no longer be (fatalistically) reasonable. Imagine that $\mathcal{USR}$ anticipates* 90% *chance that a wins where a has a large margin over b, and a* 10% *chance that b wins with a narrow margin over a. Then a rational user would incentivized to vote $b > a$ so that she is coherent in the* 10% *of cases that the reward and penalty for this pair are large, even though a is more likely to win overall.*

A solution to recover some form of fatalistic reasonableness (at least in the context of Kleros with its appeal system, see Section 5) is to have weightings only apply in the last appeal round. Then, if voters are certain that a case will be appealed, they have incentives equivalent to an incentive system without weighting, which we have seen is fatalistically reasonable. More generally, when evaluating how to vote in situations like that of the previous example, voters will have to judge how likely they are to be in the last round; if rational voters believe they are likely in the last round they will potentially take large deviations from honest votes. On might expect such deviations to be tempered by the expectation that large collective deviations are likely to cause a result that is appealed.

# 7 Heuristics on the choice of voting system

There is a rich literature on advantages and disadvantages of different voting systems in traditional social choice theory. In this section, we consider a few heuristics concerning how different voting systems might compare in the context of Schelling games.

## 7.1 Condorcet voting systems versus non-Condorcet systems

We will mostly consider Condorcet voting systems. Our reasoning for this is that the basic principle that if one candidate can beat each other candidate head-to-head, that candidate should win is easy for voters to understand. Hence, it is easier for voters to simulate how each of the other voters will think about the rules and a Schelling point is more easily formed. This is likely true even in comparison to non-Condorcet systems with relatively straightforward rules such as Instant-Runoff, where voters would nonetheless have to imagine the progression through rounds of voting.

Hence Condorcet systems should work well as long as voters *expect* there to often be Condorcet winners. Note that this is a lower and more realistic threshold than to have voters expect there to often be plurality winners with majority support. (Compare to our discussion in the introduction where we argued that a plurality voting system was problematic for a Schelling game when one does not expect there to often be plurality winners with majority support, see Figure 2.) On the other hand, if voters think there will be often situations where there is no Condorcet winner, attempting to simulate how the

behavior of other voters will affect a given system's process for resolving the ensuing Condorcet paradoxes may be challenging and make it more difficult for a Schelling point to form. For applications to Kleros, we expect to monitor the rate at which Condorcet winners are chosen.

As an added advantage, Condorcet systems appear to have less attack surface than common non-Condorcet voting systems [22]. This is particularly relevant in the context of permissionless blockchain based applications where there are few external, practical barriers to attack. Furthermore, one of the most pathological implications of the Condorcet criterion, that it implies the no-show paradox [14], is not relevant in the context of Schelling games; indeed, by making the deposit of voters sufficiently large, one can design the system so that not voting results in a larger penalty than submitting the most penalized vote.

## 7.2 Resistance to attacks in different Condorcet systems

In this section we will consider resilience of different Condorcet systems to attacks. Notice that these ideas are related to work that examines vulnerability of voting systems to manipulation and bribery in the sense of [9], [12], often looking at whether it is computationally tractable for an attacker to determine whether she can influence the outcome of the election if she controls $k$ votes. In another direction, work such as [23] looks at the probability that random voting profiles are manipulable (subject to a strategic vote) in terms of the number of manipulators.

Recall that in Section 7.1, we argued that using Condorcet voting rules is appropriate for Schelling games if voters *expect* there to often be Condorcet winners. Then, in any Condorcet system, if the "honest" answer is, in fact, for there to be a Condorcet winner $a$, for an attacker to change the outcome, they need to flip (via bribes, etc) at least one pair to produce $b > a$ where normally one would have $a > b$. Suppose that the least expensive attack strategy, in terms of paying bribes, that is capable of reversing an outcome involving $a$ costs $K$. Then, if following this strategy results in producing $b > a$ and this is sufficient for $b$ to become a Condorcet winner, there is an attack of cost $K$ that changes the outcome, regardless of which Condorcet voting system is used.

On the other hand, this cheapest attack that causes $a$ to no longer be a Condorcet winner may just create a Condorcet paradox in which $a$ is nonetheless selected by the voting system. It is natural to ask if there are Condorcet voting systems that are particularly resilient to attacks in this way. In general, we are interested in maximizing the expense of the cheapest viable attack. More formally, if we take $l_1, l_2, ..., l_{\binom{\#A}{2}}$, we define

$$k_{l_1,...,l_{\binom{\#A}{2}}} = \min_{\substack{\text{possible votes such that} \\ i\text{th pair has difference } \geq l_i \forall i}} \{\text{number of votes that need to be changed to change outcome}\}. \tag{1}$$

Then we would want to choose a voting system that, as much as possible, maximizes $k_{l_1,...,l_{\binom{\#A}{2}}}$ over different choices of the $l_j$ (here the $l_j$ act as a sort of normalization on the strength of a given victory).

Promising candidates (at least within our framework/heuristic where voters choose their vote under the *expectation* that there will usually be be a Condorcet winner) are voting systems that consider to what degree a result "deviates" from having a Condorcet winner and chose the candidate that has the smallest "deviation". Examples include Dodgson [7] and Ranked Pairs. In these systems, if an attacker that just barely flips a pair to transition from having a Condorcet winner $a$ to a Condorcet paradox, $a$ will often still exhibit the least deviation from having a Condorcet winner and be chosen as the result. In contrast, in Condorcet systems such as Alternative Smith which essentially use a second voting system as a tie-break in the event of a Condorcet paradox, $a$ may or may not be be particularly favored by the tie-break. Then in the worst case, the cheapest viable attacks will only cost the amount $K$ in bribes necessary to force a Condorcet paradox.

We illustrate these ideas:

**Example 2.** *In the Ranked Pairs system, if the attacker manages to change enough votes so that this pair just barely flips, then it will likely be (one of) the lowest ranked pair(s), hence in the Ranked Pairs system it will be disregarded. Then if there was "supposed" to be a Condorcet winner, an attack requires getting not only enough votes to flip the single pair $a > b$ to $b > a$, but the attacker must have*

- *enough votes so that $b > a$ is higher ranked than other pairs or*
- *weaken other pairs so that $b > a$ can overtake them.*

*Under the payoff system 6, weakening pairs other than that between $a$ and $b$ only has a cost to the attacker to the degree that they believed one of the outcomes in those pairs had a positive probability of winning.*

*Imagine that we have possible outcomes $a$, $b$, $c$, where in the absence of the attacker, $a$ would be the Condorcet winner. Further suppose that, in the absence of the attacker, $a > c$ by a margin of $y$ votes, $b > c$ by a margin of $x$ votes, and $a > b$ by a margin of $z$ votes, $z > x > y$.*

*Then, the minimum number of attacker votes (through bribes, etc) required to change the outcome in any Condorcet system is $y$ as this is what is required for $c > a$ to win narrowly. In order to change the outcome under Ranked Pairs, the attacker can try to weaken $b > c$ so that it becomes the weakest pair, which further requires votes such that $c > b$. By controlling $\frac{x+y}{2}$ attacker votes to vote $c > a > b$, the attacker can produce an outcome of $c$. Note $\frac{x+y}{2} > y$ as $x > y$ by assumption.*

*On the other hand under Alternative Smith, if the attacker has only $y$ votes of $c > a > b$, then she forces a runoff round where the candidate with the least number of first place votes is eliminated. If $a$ is viewed by all honest voters as an overwhelming favorite, it is likely that $b$ would not have had many first place votes and hence be eliminated, leaving $c$ to defeat $a$ as $c > a$ after the attacker votes are considered. So the number of votes the attacker needs to control is reduced in this system.*

We note already that Dodgson - where one considers the number of pairwise flips necessary for a candidate to become a Condorcet winner - and Ranked Pairs - which looks at the weakest pairwise results between candidate that need to be reversed for the graph of pairwise results to be that of having a Condorcet winner - measure deviation in different non-comparable ways, and it is not clear which produces better worst case guarantees. Dodgson has the interesting property that the Dodgson score, in that measures how many pairs must be reversed for each choice to become a Condorcet winner, provides a rough measure of how expensive it would be to bribe enough voters to make alternative choices Condorcet winners. (However, this measure only ultimately gets at an upper bound on the cost of attacks as an attacker need only change enough votes for an alternative choice to win in a Condorcet paradox; causing that choice to become a Condorcet winner is not necessary.) On the other hand, computing the winner in Dodgson is NP-hard [3], so it is likely not a viable choice for our applications regardless.

One might ask more generally if this desired property of maximizing resistance to the cheapest attacks as formulated in equation 1 can be encoded as a distance with a corresponding distance rationalizable voting system in the sense of [8] that has optimal worst case resistance to attacks. Note that even if such a system exists, it would not necessarily be guaranteed to have other desirable properties (such as monotonicity or being able to be computed in polynomial time). This is a potential subject of future work.

# 8 Observations

## 8.1 Incentive compatible systems can exist

**Example 3.** *Consider the following (simplistic) system: users rank all possible outcomes creating an ordering in $L(A)$ and place a deposit $D$. The winning outcome is decided by plurality voting and any voter who did not correctly place the winning outcome first loses their deposit (in this example, we avoid considering how the deposits would be redistributed, so these lost deposits are burnt). Then for a given $v = vote_{\mathcal{USR}} = a_1 > a_2 > \ldots$, the user's expected return is*

$$-D \cdot (1 - prob(a_1 \ wins : \mathcal{USR} \ votes \ v)).$$

*Here $prob(a_i \ wins : \mathcal{USR} \ votes \ v)$ represents the probability that the winning choice is $a_i$ if $\mathcal{USR}$ casts the vote $v$, based on $\mathcal{USR}$'s priors regarding all other votes $Vot(\mathcal{USR})$ and the chances of appeal. Note that this is optimized by maximizing $prob(a_1 \ wins : \mathcal{USR} \ votes \ v)$. We claim that it is always in the voter's interest to produce a vote $v$ such that*

$$prob(a_1 \ wins : \mathcal{USR} \ votes \ v) \geq prob(a_i \ wins : \mathcal{USR} \ votes \ v) \tag{2}$$

*for all $a_i$. If this were not the case, then by voting $v_2 = a_i > a_1 > a_2 > \ldots > a_{i-1} > a_{i+1} > \ldots$, the voter would have*

$$prob(a_i \ wins : \mathcal{USR} \ votes \ v_2) \geq prob(a_i \ wins : \mathcal{USR} \ votes \ v) > prob(a_1 \ wins : \mathcal{USR} \ votes \ v)$$

*as plurality is a monotonic voting system (for any given possibility in $Vot(\mathcal{USR})$ of how the other voters vote, ranking $a_i$ higher cannot cause it to lose when $\mathcal{USR}$ votes $v_2$ situations it would have won if $\mathcal{USR}$ votes $v$). Hence, whichever $a_i$ that produces the largest value of prob($a_i$ wins : modified $v$ to move $a_i$ first), both maximizes the payoff and satisfies inequality 2.*

*Then, as in this simplistic system, only the first choice vote of a user has any effect on either the winner or the payout, $\mathcal{USR}$ does no worse by choosing $a_2, a_3, \ldots$ such that prob($a_i$ wins : $\mathcal{USR}$ votes $v$) $\geq$ prob($a_{i+1}$ wins : $\mathcal{USR}$ votes $v$) for all $i$.*

## 8.2 Impossibility results

**Proposition 4.** *Suppose $\#A = 3$. Then there does not exist any incentive compatible voting system, payoff system pair such that*

- *the voting system is monotonic*
- *there exists some choice of voters for voters other than $\mathcal{USR}$ such that the voting system produces a constant result regardless of how $\mathcal{USR}$ votes*
- *the payoff system is strictly reasonable*
- *both the voting system and the payoff system are symmetric in the options (i.e. a permutation of the outcomes for all voters results in correspondingly permuted results and preserves payouts - on the level of the voting system this is the property of neutrality).*

**Summary of proof** The Gibbard-Satterthwaite Theorem gives some tactical vote. There are a number of cases for what form this vote takes (i.e. what rearrangement of the voter's honest preferences resulting in which winner). Then the basic idea is that, there can arise situations where by employing the "tactical" vote, the voter increase the chances of an outcome that she ranks highly in exchange for the probability of her $n$th choice being lower than that of her $n+1$st choice. The strictly reasonable assumption implies that, if the probabilities of the $n$th and $n+1$st choices are close enough under the tactical vote, this dishonest vote will result in a higher payout. For example, if Gibbard-Satterthwaite implies the existence of a scenario where $\mathcal{USR}$ voting $a > b > c$ results in $b$ winning, and voting $a > c > b$ results in $a$, then the voter might be able to transform a situation where she thinks $a$ has a 51% chance, $b$ has a 49% chance and $c$ has a 1% chance if she votes $a > b > c$ into a situation where $a$ has a 97% chance, $c$ has a 1% chance and $b$ has a 2% chance if she votes $a > c > b$. The proof then shows that there exists some scenario like this for each possible case.

Note that the assumption of monotonicity is to reduce the number of cases that need to be considered. One might hope that this condition can be removed, either by finding a better proof technique or via a computer assisted proof.

*Proof.* Note that the condition that there exists some scenario of other user's votes where the voting system produces a constant result regardless of $\mathcal{USR}$'s vote implies that the voting system is non-dictatorial.

Let $\{a, b, c\}$ be the three possible voting outcomes. By the assumed symmetry properties, the voting system must be surjective. Then, as the voting system is not dictatorial, by the Gibbard-Satterthwaite Theorem, there exists some $v_{-i}$, $r_i$, $s_i$ such that $f(r_i, v_{-i}) <_{r_i} f(s_i, v_{-i})$.

Without loss of generality, take $r_i$ to be the vote $a > b > c$. Then as $f(r_i, v_{-i}) <_{r_i} f(s_i, v_{-i})$, we must have $f(r_i, v_{-i}) = b$ or $f(r_i, v_{-i}) = c$.

**Case 1:** $f(r_i, v_{-i}) = b$ In this case we must have $f(s_i, v_{-i}) = a$. Note by monotonicity, if $\mathcal{USR}$ votes $b > a > c$ or $b > c > a$ the outcome cannot be $a$. Furthermore, if voting $c > a > b$ or $c > b > a$ results in an outcome of $a$, so does voting $a > c > b$. Hence we can assume that $s_i$ is $a > c > b$.

Consider 4 to be the scenario of this tactical vote, while 1, 2, and 3 are the scenarios that $a$, $b$, and $c$ win respectively regardless of $\mathcal{USR}'s$ vote (note that such a scenario is guaranteed for one of the three outcomes by assumption; hence it must exist for all three by the symmetry assumptions).

By symmetry

$$G(r_i, v_{-i} = 1) = G(s_i, v_{-i} = 1)$$

$$G(r_i, v_{-i} = 2) = G(s_i, v_{-i} = 3)$$

$$G(r_i, v_{-i} = 3) = G(s_i, v_{-i} = 2)$$

By Lemma 1

$$G(s_i, v_{-i} = 4) - G(r_i, v_{-i} = 4) > 0$$

and

$$G(r_i, v_{-i} = 2) - G(s_i, v_{-i} = 2) > 0$$

Denote by $p_j$ the probability that $\mathcal{USR}$ assigns to the $j$th scenario occurring, $p_1 + p_2 + p_3 + p_4 = 1$.

Take $p_3$, $p_4 > 0$ such that $3p_3 + 2p_4 < 1$. Then denote

$$\epsilon = \frac{1}{2} \min \left\{ \frac{1 - 3p_3 - 2p_4}{2}, p_4, p_4 \cdot \frac{G(s_i, 4) - G(r_i, 4)}{G(r_i, 2) - G(s_i, 2)} \right\} > 0`.$$

Then take $p_2 = p_3 + \epsilon$ and $p_1 = 1 - p_2 - p_3 - p_4$. Note $p_2 = p_3 + \epsilon < p_3 + p_4 < 1$, so $p_2 \in (0, 1)$. Similarly,

$$p_1 = 1 - p_2 - p_3 - p_4 = 1 - 2p_3 - p_4 - \epsilon > p_3 + p_4 + \epsilon > 0$$

because we have $\frac{1-3p_3-2p_4}{2} > \epsilon$ by construction. So $p_1 \in (0, 1)$ as well.

Then we have

$$E[G(s_i, v_{-i})] - E[G(r_i, v_{-i})] = \sum_{j=1}^{4} p_j \cdot (G(s_i, v_{-i} = j) - G(r_i, v_{-i} = j))$$

$$= (p_2 - p_3)\,(G(s_i, v_{-i} = 2) - G(r_i, v_{-i} = 2)) + p_4\,(G(s_i, v_{-i} = 4) - G(r_i, v_{-i} = 4))$$

$$= \epsilon\left(G(s_i, v_{-i} = 2) - G(r_i, v_{-i} = 2)\right) + p_4\left(G(s_i, v_{-i} = 4) - G(r_i, v_{-i} = 4)\right) > 0,$$

where we have used the implications of symmetry on the payoffs and the definitions of $p_2$ and $\epsilon$.

Namely, $s_i$ gives a higher expected payoff than $r_i$, however $s_i$ is not honest as its third choice has a higher probability of occurring than its second ($p_3 + \epsilon$ versus $p_3$).

Similarly, we see that, (as we have seen $p_1 > p_3 + p_4 + \epsilon$, $p_2 = p_3 + \epsilon$, and $p_4 > \epsilon$) that all other possible votes are dishonest, even allowing for the different possibilities for how those votes can influence scenario 4.

| $\mathcal{USR}$ vote incl. effect on 4 \vote outcome by scenario | a | b | c |
|---|---|---|---|
| $b > a > c \to b$ | 1 | 2,4 | 3 |
| $b > c > a \to b$ | 1 | 2,4 | 3 |
| $b > a > c \to c$ | 1 | 2 | 3,4 |
| $b > c > a \to c$ | 1 | 2 | 3,4 |
| $c > a > b \to a$ | 1,4 | 2 | 3 |
| $c > b > a \to a$ | 1,4 | 2 | 3 |
| $c > a > b \to b$ | 1 | 2,4 | 3 |
| $c > b > a \to b$ | 1 | 2,4 | 3 |
| $c > a > b \to c$ | 1 | 2 | 3,4 |
| $c > b > a \to c$ | 1 | 2 | 3,4 |

**Case 2:** $f(r_i, v_{-i}) = C$ In this case we could have either $f(s_i, v_{-i}) = a$ or $f(s_i, v_{-i}) = b$.

Similar to case 1, we have scenarios 1, 2, and 3 where $a$, $b$, and $c$ win respectively regardless of the vote of $\mathcal{USR}$. Also, we have scenario 4, which is the scenario in which we have the tactical vote given by Gibbard-Satterthwaite. Note by monotonicity, as $a > b > c \to c$, we must have $a > c > b \to c$, and $c > a > b \to c$. Furthermore, as $a > b > c \to c$, we cannot have $b > a > c \to a$ or $b > c > a \to a$. However, we also cannot have $b > a > c \to c$ as this would imply $c > b > a \to c$ and would prevent $b > c > a \to b$. This would imply that all votes result in $c$, contradicting the existence of $s_i$. Hence $b > a > c \to b$.

Additionally, for each of the six permutations of $\{a, b, c\}$, we have a corresponding tactical vote, as given in the following table:

| $\mathcal{USR}$ vote \scenario | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|
| $a > b > c$ | $c$ | $b$ | $a$ or $b$ | $a$ or $b$ | $a$ | $a$ |
| $a > c > b$ | $c$ | $b$ | $a$ | $a$ | $a$ or $c$ | $a$ or $c$ |
| $b > a > c$ | $b$ | $b$ | $a$ or $b$ | $a$ or $b$ | $c$ | $a$ |
| $b > c > a$ | $b$ or $c$ | $b$ or $c$ | $b$ | $b$ | $c$ | $a$ |
| $c > a > b$ | $c$ | $c$ | $a$ | $b$ | $a$ or $c$ | $a$ or $c$ |
| $c > b > a$ | $b$ or $c$ | $b$ or $c$ | $a$ | $b$ | $c$ | $c$ |

For the entries in this table that are not yet determined, we have several subcases (eliminating some possibilities again by monotonicity). In each case we will take $r_i$ to be the vote $a > b > c$ and $s_i$ to be $b > a > c$.

**Subcase $b > c > a \to b$ and $c > b > a \to b$ in scenario 4**

Note that

$$G(a>b>c,7)=G(a>b>c,9)=G(b>a>c,4)=G(b>a>c,5)>G(b>a>c,6)$$

$$>G(a>b>c,5)=G(a>b>c,6)=G(b>a>c,9)=G(b>a>c,7)>G(a>b>c,4)$$

by various applications of symmetry and (strict) reasonableness.

We will have $p_1=p_2+p_3=p_8=0$, $p_4=p_5$, $p_7=p_9$.

Then

$$E[G(s_i,v_{-i})]-E[G(r_i,v_{-i})]$$

$$=p_4(G(a>b>c,7)-G(a>b>c,4))+(2p_5-2p_7)(G(a>b>c,7)-G(a>b>c,5)).$$

Take

$$p_4=\frac{1}{2}\left[\frac{1/4}{1+\frac{3/2+\frac{1}{2\frac{G(a>b>c,7)-G(a>b>c,5)}{G(a>b>c,7)-G(a>b>c,4)}}}{4}}\right]+\frac{1}{2}\left[\frac{1}{3/2+\frac{1}{2\frac{G(a>b>c,7)-G(a>b>c,5)}{G(a>b>c,7)-G(a>b>c,4)}}}\right].$$

Then as $p_4<\frac{1}{3/2+\frac{1}{2\frac{G(a>b>c,7)-G(a>b>c,5)}{G(a>b>c,7)-G(a>b>c,4)}}}$, we can take

$$p_5=\frac{1-p_4\left(\frac{3}{2}+\frac{1}{2\frac{G(a>b>c,7)-G(a>b>c,5)}{G(a>b>c,7)-G(a>b>c,4)}}\right)}{4}>0.$$

As $p_4>\frac{1/4}{1+\frac{3/2+\frac{1}{2\frac{G(a>b>c,7)-G(a>b>c,5)}{G(a>b>c,7)-G(a>b>c,4)}}}{4}}$, we have $p_4>p_5$.

Furthermore, take

$$p_7=\frac{1}{2}\left[\frac{p_4}{2\frac{G(a>b>c,7)-G(a>b>c,5)}{G(a>b>c,7)-G(a>b>c,4)}}+p_5\right]+\frac{1}{2}\left[\frac{p_4}{2}+p_5\right]=\frac{p_4}{4}\left[\frac{1}{\frac{G(a>b>c,7)-G(a>b>c,5)}{G(a>b>c,7)-G(a>b>c,4)}}+1\right]+p_5>p_5.$$

Note that $p_4+2p_5+2p_7=1$ and all $p_j$ are non-negative, so all $p_j\in[0,1]$.

Then, as $G(a>b>c,7)-G(a>b>c,5)<G(a>b>c,7)-G(a>b>c,4)$, we have

$$\frac{p_4}{2}+p_5<p_7<\frac{p_4}{2\frac{G(a>b>c,7)-G(a>b>c,5)}{G(a>b>c,7)-G(a>b>c,4)}}+p_5$$

Then we have that

- $p_4>2\left(\frac{G(a>b>c,7)-G(a>b>c,5)}{G(a>b>c,7)-G(a>b>c,4)}\right)(p_7-p_5)$
- $2p_7>p_4+2p_5$
- $p_4>p_5$
- $p_7>p_5$

- $p_5 = p_6$, $p_7 = p_9$, $p_1 = p_2 = p_3 = p_8 = 0$, $\sum_{j=1}^{9} p_j = 1$, $p_j \in [0,1]\ \forall j$.

The first condition tells us that $b > a > c$ has a higher payoff than $a > b > c$. Using Table 8.2 the other conditions tell us that all votes except for (potentially) $a > b > c$ are dishonest.

**Subcase $b > c > a \to b$ and $c > b > a \to c$ in scenario 4**

Here, we will be able to choose $p_j$ that illustrate an unavoidable non-incentive compatible situation where $p_3 = p_5 = p_6 = p_7 = p_8 = 0$.

Note that

$$G(a > b > c, 9) = G(c > a > b, 4) = G(b > a > c, 4)$$

$$> G(b > a > c, 9) = G(a > c > b, 4) > G(a > b > c, 4)$$

by successive applications of symmetry and (strict) reasonableness. By a similar analysis applied to the payoffs of the constant scenarios,

$$G(a > b > c, 1) = G(b > a > c, 2) > G(a, b, c, 2) = G(b > a > c, 1).$$

Then,

$$E[G(s_i, v_{-i})] = p_1 G(b > a > c, 1) + p_2 G(b > a > c, 2) + p_4 G(b > a > c, 4) + p_9 G(b > a > c, 9)$$

$$E[G(r_i, v_{-i})] = p_1 G(a > b > c, 1) + p_2 G(a > b > c, 2) + p_4 G(a > b > c, 4) + p_9 G(a > b > c, 9)$$

Then

$$E[G(s_i, v_{-i})] - E[G(r_i, v_{-i})]$$

$$= p_1(G(a > b > c, 2) - G(a > b > c, 1)) + p_2(G(a > b > c, 1) - G(a > b > c, 2))$$

$$+ p_4(G(a > b > c, 9) - G(a > b > c, 4)) + p_9(G(b > a > c, 9) - G(a > b > c, 9)).$$

So

$$E[G(s_i, v_{-i})] - E[G(r_i, v_{-i})] > 0$$

$$\Leftrightarrow p_9 < p_4 \frac{G(a > b > c, 9) - G(a > b > c, 4)}{G(a > b > c, 9) - G(b > a > c, 9)} + (p_2 - p_1) \frac{G(a > b > c, 1) - G(a > b > c, 2)}{G(a > b > c, 9) - G(b > a > c, 9)}$$

Then, including observations we make from Table 8.2, we have that $s_i$ has a greater payout than $r_i$ and all votes other than (potentially) $r_i$ are dishonest if

- $p_9 < p_4 \frac{G(a>b>c,9) - G(a>b>c,4)}{G(a>b>c,9) - G(b>a>c,9)} + (p_2 - p_1) \frac{G(a>b>c,1) - G(a>b>c,2)}{G(a>b>c,9) - G(b>a>c,9)}$
- $p_1 + p_9 > p_2 + p_4$
- $p_2 > p_4 + p_9$
- $p_1 + p_9 > 0$
- $p_1, p_2, p_4, p_9 \in [0,1]$, $p_1 + p_2 + p_4 + p_9 = 1$.

**Subsubcase:** $\frac{G(a>b>c,1)-G(a>b>c,2)}{G(a>b>c,9)-G(b>a>c,9)} > 1$

Here we take $p_1 = 1/3$, $p_2 = p_1 + \epsilon$, $p_4 = 0$, $p_9 = \frac{\epsilon\frac{G(a>b>c,1)-G(a>b>c,2)}{G(a>b>c,9)-G(b>a>c,9)}+\epsilon}{2}$, where $\epsilon$ is chosen so that $p_1 + p_2 + p_9 = 1$. As $\frac{G(a>b>c,1)-G(a>b>c,2)}{G(a>b>c,9)-G(b>a>c,9)} > 1$, $p_9$ is such that the first two conditions are satisfied. As all $p_j$ are clearly non-negative, and they add to 1 by construction, they are all in $[0, 1]$. Moreover, as $p_2 > p_1 = 1/3$, $p_4 + p_9 < p_2$.

**Subsubcase:** $\frac{G(a>b>c,1)-G(a>b>c,2)}{G(a>b>c,9)-G(b>a>c,9)} = 1$

Here we take $p_1 = 1/3$, $p_2 = p_1 + \epsilon$, $p_4 = \frac{\frac{1}{3}-2\epsilon}{\frac{3}{2}+\frac{G(a>b>c,9)-G(a>b>c,4)}{2(G(a>b>c,9)-G(b>a>c,9))}}$, and

$$p_9 = \frac{1}{2}\left(p_4\frac{G(a>b>c,9)-G(a>b>c,4)}{G(a>b>c,9)-G(b>a>c,9)}+\epsilon\right)+\frac{1}{2}(p_4+\epsilon)$$

$$= \epsilon + p_4\left(\frac{1}{2}+\frac{G(a>b>c,9)-G(a>b>c,4)}{2(G(a>b>c,9)-G(b>a>c,9))}\right)$$

$$= \epsilon+\frac{\frac{1}{3}-2\epsilon}{\frac{3}{2}+\frac{G(a>b>c,9)-G(a>b>c,4)}{2(G(a>b>c,9)-G(b>a>c,9))}}\cdot\left(\frac{1}{2}+\frac{G(a>b>c,9)-G(a>b>c,4)}{2(G(a>b>c,9)-G(b>a>c,9))}\right).$$

Due to the first equality in the choice of $p_9$ and using $\frac{G(a>b>c,1)-G(a>b>c,2)}{G(a>b>c,9)-G(b>a>c,9)} = 1$, we have the first two required constraints. Then the definition of $p_4$ is such that $p_1 + p_2 + p_4 + p_9 = 1$ by construction. As $p_2 > p_1 = 1/3 > p_4 + p_9$, we have the third constraint. Then we can clearly take $\epsilon > 0$ small enough so that all of the $p_j$ are in $[0, 1]$. Taking such a choice of $\epsilon$ completes this case.

**Subsubcase:** $\frac{G(a>b>c,1)-G(a>b>c,2)}{G(a>b>c,9)-G(b>a>c,9)} < 1$ We will take $p_2 = p_1 + \epsilon$ for $\epsilon > 0$. Take $p_4 = 2\epsilon\frac{1-\frac{G(a>b>c,1)-G(a>b>c,2)}{G(a>b>c,9)-G(b>a>c,9)}}{\frac{G(a>b>c,9)-G(a>b>c,4)}{G(a>b>c,9)-G(a>b>c,4)}-1}$. Note $p_4 > 0$ by our observations and assumptions on the payouts. Moreover, note that this choice of $p_4$ implies

$$\epsilon+p_4 < p_4\frac{G(a>b>c,9)-G(a>b>c,4)}{G(a>b>c,9)-G(a>b>c,4)}+\epsilon\frac{G(a>b>c,1)-G(a>b>c,2)}{G(a>b>c,9)-G(b>a>c,9)}.$$

Take

$$p_9 = \frac{1}{2}(\epsilon+p_4)+\frac{1}{2}\left(p_4\frac{G(a>b>c,9)-G(a>b>c,4)}{G(a>b>c,9)-G(a>b>c,4)}+\epsilon\frac{G(a>b>c,1)-G(a>b>c,2)}{G(a>b>c,9)-G(b>a>c,9)}\right).$$

Then, these choices, with $\epsilon = p_2 - p_1$, imply the first two constraints.

Then, for a given $\epsilon$, choose $p_1$ such that $p_1 + p_2 + p_4 + p_9 = 1$. Note as the choice of $\epsilon$ tends to zero, $p_4$ and $p_9$ also tend to zero. Hence, in particular, $\epsilon$ can be chosen small enough so that $p_4 + p_9 < 1/3$. Then $p_2 = p_1 + \epsilon > p_1 > p_1$, $p_1 + p_2 + p_4 + p_9 = 1$ imply that $p_2 > \frac{1}{2}(1 - p_4 - p_9) > \frac{1}{3}$, giving the third constraint. Finally, for $\epsilon$ small enough such that $p_4 + p_9 < 1/3$, $p_1$ is chosen as $p_1 = \frac{1-p_4-p_9-\epsilon}{2} > \frac{1}{3} - \frac{\epsilon}{2}$. Then $p_2 > \frac{1}{3} + \frac{\epsilon}{2}$. Hence, for sufficiently small $\epsilon$, all $p_j$ are non-negative, and as they sum to 1 they are all in $[0, 1]$.

**Subcase $b > c > a \to c$ and $c > b > a \to c$ in scenario 4**

Note that $G(b > a > c, 4) = G(c > a > b, 5) = G(b > a > c, 5) > G(a > b > c, 5) = G(a > c > b, 4)$ by successive applications of symmetry and (strict) reasonableness. Similarly, we find

$$G(b > a > c, 4) = G(b > a > c, 5) = G(a > b > c, 6) = G(a > b > c, 9)$$

$$> G(a > b > c, 5) = G(b > a > c, 6) = G(b > a > c, 6) > G(a > b > c, 4)$$

Take $p_1 = p_2 = p_3 = p_7 = p_8 = 0$, $p_5 = p_6 = 1/4$. Take

$$p_9 = \frac{1}{3 + \frac{G(s_i,4) - G(s_i,9)}{G(s_i,4) - G(r_i,4)}}$$

and $p_4 = 1/2 - p_9$.

Then as $\frac{G(s_i,4) - G(s_i,9)}{G(s_i,4) - G(r_i,4)} \in (0, 1)$, we have that all $p_j \in [0, 1]$. Note $\sum_{j=1}^{9} p_j = 1$. Then

$$E[G(s_i, v_{-i})] - E[G(r_i, v_{-i})]$$

$$= p_4[G(s_i,4) - G(r_i,4)] + p_5[G(s_i,5) - G(r_i,5)] + p_6[G(s_i,6) - G(r_i,6)] + p_9[G(s_i,9) - G(r_i,9)]$$

$$= (p_4 - p_9)p_5[G(s_i,4) - G(s_i,9)] + p_4[G(s_i,9) - G(r_i,4)] > 0$$

as

$$p_4 = \frac{1}{2} - \frac{1}{3 + \frac{G(s_i,4) - G(s_i,9)}{G(s_i,4) - G(r_i,4)}}$$

$$\Rightarrow p_4 > p_9 \frac{G(s_i,4) - G(s_i,9)}{G(s_i,4) - G(r_i,4)}.$$

Furthermore, as we have $p_4 < \min\{p_5, p_6, p_9\}$ and considering the various cases in Table 8.2, we see that all votes other than $a > b > c$ and $c > b > a$ are dishonest. Moreover, as $p_4 < p_9$, by reasonableness the vote $c > b > a$ has a worse payoff than $a > b > c$. Hence, the vote that maximizes payoffs is not honest in this case.

| Table 8.2 | | | |
|---|---|---|---|
| $\mathcal{USR}$ vote \vote outcome by scenario | a | b | c |
| $a > c > b$ assuming $b > c > a \to b$ and $c > b > a \to b$ in scenario 4 | 1,6,7,8 | 2,5 | 3,4,9 |
| $b > a > c$ assuming $b > c > a \to b$ and $c > b > a \to b$ in scenario 4 | 1,7,9 | 2,4,5,6 | 3,8 |
| $b > c > a$ assuming $b > c > a \to b$ and $c > b > a \to b$ in scenario 4 | 1,9 | 2,4,6,7 | 3,5,8 |
| $c > a > b$ assuming $b > c > a \to b$ and $c > b > a \to b$ in scenario 4 | 1,6,8 | 2,7 | 3,4,5,9 |
| $c > b > a$ assuming $b > c > a \to b$ and $c > b > a \to b$ in scenario 4 | 1,6 | 2,5,7 | 3,4,8,9 |
| $a > c > b$ assuming $b > c > a \to b$ and $c > b > a \to c$ in scenario 4 | 1,6,7,8,9 | 2,5 | 3,4 |
| $b > a > c$ assuming $b > c > a \to b$ and $c > b > a \to c$ in scenario 4 | 1,9 | 2,4,5,6,7 | 3,8 |
| $b > c > a$ assuming $b > c > a \to b$ and $c > b > a \to c$ in scenario 4 | 1,9 | 2,4,5,6,7 | 3,8 |
| $c > a > b$ assuming $b > c > a \to b$ and $c > b > a \to c$ in scenario 4 | 1,6 | 2,7 | 3,4,5,8,9 |
| $c > b > a$ assuming $b > c > a \to b$ and $c > b > a \to c$ in scenario 4 | 1,6 | 2,7 | 3,4,5,8,9 |
| $a > c > b$ assuming $b > c > a \to c$ and $c > b > a \to c$ in scenario 4 | 1,6,7,9 | 2,5 | 3,4,8 |
| $b > a > c$ assuming $b > c > a \to c$ and $c > b > a \to c$ in scenario 4 | 1,6,9 | 2,4,5,7 | 3,8 |
| $b > c > a$ assuming $b > c > a \to c$ and $c > b > a \to c$ in scenario 4 | 1,9 | 2,5,6,7 | 3,4,8 |
| $c > a > b$ assuming $b > c > a \to c$ and $c > b > a \to c$ in scenario 4 | 1,6,9 | 2,7 | 3,4,5,8 |
| $c > b > a$ assuming $b > c > a \to c$ and $c > b > a \to c$ in scenario 4 | 1,6 | 2,5,7 | 3,4,8,9 |
| | | | |

□

**Theorem 3.** *Suppose* $\#A \geq 3$. *Then there does not exist any incentive compatible voting system, payoff system pair such that*

- *the voting system is monotonic*
- *if* $a_1, \ldots, a_k = S \cup T$, *where* $S$ *and* $T$ *are disjoint, non-empty sets, and if all voters other than* $\mathcal{USR}$ *vote all elements in* $S$ *ahead of all elements in* $T$, *then the result is in* $S$ *regardless of the vote of* $\mathcal{USR}$
- *the payoff system is strictly reasonable*

- *both the voting system and the payoff system are symmetric in the options (i.e. a permutation of the outcomes for all voters results in correspondingly permuted results and preserves payouts on the level of the voting system this is the property of neutrality).*

*Proof.* Note that the second condition implies that if all voters other than $\mathcal{USR}$ vote a choice first, that choice wins. Then this implies that there exists some $v_{-i}$ that produces a constant result regardless of the vote of $\mathcal{USR}$. Hence, Proposition 4 gives the result for the case $\#A = 3$.

In general, suppose $\#A = k > 3$. Consider the subset of possible votes where $\mathcal{USR}$ and all other voters rank some permutation of $\{a, b, c\}$ as their first three votes and rank $a_4 > a_5 > \ldots > a_k$ as the remainder of their lists. Note that by the second assumption of the theorem, either $a$, $b$, or $c$ must win in any such scenario. This is (isomorphic to) a voting rule in which voters only rank the three choices $a$, $b$, and $c$, so by the Gibbard-Satterthwaite Theorem, their exists some $r_i$, $s_i$, $v_{-i}$ where all votes are of this form such that $f(r_i, v_{-i}) <_{r_i} f(s_i, v_{-i})$. Without loss of generality, suppose $r_i$ is the vote $a > b > c > a_4 > \ldots > a_k$. Then, similar to the proof of Proposition 4, take scenarios 4 through 9 to be the various permutations of $v_{-i}$. Scenarios 1, 2, and 3 are taken to be those where $a$, $b$, and $c$ respectively win regardless of $\mathcal{USR}$'s vote (which exist by our second and fourth assumptions). Furthermore, we take weightings in the $k$th case are now chosen according to the equations that define the $p_j$ in Proposition 4 where, when these choices depend on $G(\text{vote}, \text{scenario})$, $a_4, \ldots, a_k$ are appended to the end of the vote.

In the proof of Proposition 4, (in each subcase) we considered a table of all possible votes and we see that the choice we made for $p_j$ imply that there is always some dishonest vote that has a higher payoff than $r_i$, and that all votes are either dishonest or have a payoff upper bounded by that of $r_i$. We make the same argument in this case.

Here we have the same (sub)cases, appending $a_4 > \ldots > a_k$ where required. We can take tables of possible $\mathcal{USR}$ votes (restricting for the moment to votes of the form of a permutation of $\{a, b, c\}$ followed by $a_4 > \ldots > a_k$) that have the same structure because we can make the same arguments by monotonicity and symmetry, and to the degree that we used process of elimination in Proposition 4 to assign results, this is still valid as the only possible results are still $a, b, c$ (even though we are not assuming that a vote of $a > b > c > d$ necessarily produces the same result as $a > b > c$ did in the $\#A = 3$ case).

Then the inequalities we consider on the $p_j$ showing these votes as dishonest also still hold. For the cases in which we saw that a vote was (potentially) honest but had a payoff bounded by that of $a > b > c$ by reasonableness, again, the inequalities we have on the $p_j$ show that the vote of the $k$ choices that places $a, b, c$ in this order followed by $a_4 > a_5 > a_6 > \ldots > a_k$ has a payoff bounded by that of $a > b > c > a_4 > ... > a_k$ by reasonableness.

Hence we have that all votes of the form permutation of $\{a, b, c\}$ followed by $a_4 > \ldots > a_k$ other than $a > b > c > a_4 > ... > a_k$ are either dishonest or have a payoff bounded by that of $a > b > c > a_4 > ... > a_k$. This is not

necessarily true of votes that are not of the form permutation of $\{a, b, c\}$ followed by $a_4 > \ldots > a_k$. However, based on our choices of $p_j$ and the probabilities that these give us for each outcome winning in each case according to the tables in the proof of Proposition 4, we can take a small enough $\epsilon$ to essentially allow us to alter the $p_j$ by $\epsilon$ while preserving the orders of likelihoods of winning and orders of payoffs that we have observed:

$$\epsilon_1 = \min_{\substack{i,j=a,b, \text{ or } c, \text{ used to see} \\ \text{dishonesty for vote } x \text{ in some case,} \\ x, \text{cases}}} \{|\text{Prob}(\text{outcome } i \ : \mathcal{USR} \text{ votes } x) - \text{Prob}(\text{outcome } j \ : \mathcal{USR} \text{ votes } x)|\}$$

$$\epsilon_2 = p_6 - p_4 \qquad \begin{array}{c}(\text{ from the } b>c>a>a_4>\ldots>a_k \to C, \\ c>b>a>a_4>\ldots>a_k \to C \text{ in scenario 4 subcase})\end{array}$$

$$\epsilon_3 = \frac{\min_{\text{cases}} |E[(G(s_i, v_{-i}))] - E[(G(r_i, v_{-i}))]|}{\#\text{scenarios} \ \cdot \max_{\text{scenario}=j,\text{cases}} \{|G(s_i, v_{-i} = j) - G(r_i, v_{-i} = j)|\}}.$$

Note that the arguments we made (in Proposition 4 which we have seen here generalize) that show that various votes are dishonest or that $s_i$ offers a better payout than $r_i$, we have $\epsilon_1$, $\epsilon_2$, $\epsilon_3 > 0$. Then

$$\epsilon = \frac{1}{15} \min \{\epsilon_1, \epsilon_2, \epsilon_3\}.$$

Then we consider a slightly modified prior for $\mathcal{USR}$ as follows: for the scenarios already consider we take

$$p_j^* = (1 - 7\epsilon)p_j + \mathbf{1}_{\substack{j=1,2, \text{ or } 3, \text{ i.e. constant} \\ \text{scenarios where a, b, or c win}}} \cdot 2\epsilon,$$

and we choose $p_{a_4}^*, p_{a_5}^*, \ldots, p_{a_{k-1}}^* \in (0, 1)$ such that

$$p_{a_4}^* > p_{a_5}^* > \ldots > p_{a_{k-1}}^*$$

and

$$p_{a_4}^* + p_{a_5}^* + \ldots + p_{a_{k-1}}^* = \epsilon.$$

Hence, the sum of all $p_j^*$ and $p_{a_l}^*$'s is 1. Then we take a prior where the scenarios considered above have probability $p_j^*$ of occurring and we additionally have scenarios where each of $a_l$, $l > 3$ wins regardless of $\mathcal{USR}$'s vote with probability $p_{a_j}^*$.

Then, by our choice of $\epsilon$, we still see that all votes of form permutation of $\{a, b, c\}$ followed by $a_4 > \ldots > a_k$ other than $a > b > c > a_4 > ... > a_k$ are either dishonest or have a payoff bounded by that of $a > b > c > a_4 > ... > a_k$. Furthermore, (noting that the only scenario where an $a_l$ for $l > 3$ can win is the constant scenario where it always wins) now we can eliminate all votes not of the form permutation of $\{a, b, c\}$ followed by $a_4 > \ldots > a_k$ as dishonest.

Finally, the inequalities we have on the $p_j$ (and hence on $p_j^*$ by our choice of $\epsilon$) still show that $b > a > c > a_4 > \ldots > a_k$ or $a > c > b > a_4 > \ldots > a_k$

(depending on the case) has a better payoff than $a > b > c > a_4 > \ldots > a_k$. So we still have that there exists a dishonest vote with a higher payoff than any honest vote.

□

**Remark 1.** *Note that the $S, T$ assumption is, in a sense, a strengthened version of assuming that the system is non-dictatorial. The assumption of monotonicity is an artifact of the method of proof we used, specifically to reduce the number of cases that needed to be considered. One might hope that this condition can be removed. On the other hand the assumption of strictly reasonable is necessary as we saw in Example 3. The symmetry property is a strong version of "neutrality" that applies to the payoff systems as well as to the voting system. (Note that the voting system is not assumed to be anonymous i.e. that it is symmetric under permutations of the voters; we allow for some voters to be weighted more than other as might the case in Kleros in situations where one participant has many tokens and is drawn more than once for the same case.)*

## 8.3 For some systems it is not necessarily possible to provide an "honest" response ranked by probability

In Section 8.2 we saw that, when certain natural requirements are placed on a Schelling game voting and payoff system, there will inevitably be situations where it is in the economic interest of participants to cast dishonest votes in the sense of Definition 1. In this section we will see that for several specific voting systems, there are situations where it is impossible for a voter to cast an honest vote at all.

**Example 4.** *Consider an instant-runoff voting system that is attempting to decide between three outcomes $\{a, b, c\}$.*

*Suppose $\mathcal{USR}$ has two votes (this avoids having situations with ties). Before voting, her probability distribution for the voters of others $Vot(\mathcal{USR})$ is given as follows*

- *There is a* 49% *chance that there will be*
  - 21 *votes for $a > b > c$*
  - 22 *votes for $b > c > a$*
  - 20 *votes for $c > a > b$*
- *There is a* 49% *chance that there will be*
  - 20 *votes for $a > b > c$*
  - 21 *votes for $b > c > a$*
  - 22 *votes for $c > a > b$*
- *There is a* 2% *chance that there will be*

– 22 *votes for* $a > b > c$
– 20 *votes for* $b > c > a$
– 21 *votes for* $c > a > b$

*Then, regardless of which of the possibilities in* $Vot(\mathcal{USR})$ *occurs and regardless of how* $\mathcal{USR}$ *votes,* $a$ *wins any eventual duels against* $b$*,* $b$ *wins any eventual duels against* $c$*, and* $c$ *wins any eventual duels against* $a$*.*

*If* $\mathcal{USR}$ *ranks* $a$ *first with her two votes, then there is a* 49% *chance that the runoff will be between* $a$ *and* $b$ *and a* 51% *chance that the runoff will be between* $a$ *and* $c$*. Hence* $a$ *has a* 49% *chance of winning,* $b$ *a* 0% *chance of winning, and* $c$ *a* 51% *chance of winning.*

*If* $\mathcal{USR}$ *ranks* $b$ *first with her two votes, then there is a* 51% *chance that the runoff will be between* $a$ *and* $b$ *and a* 49% *chance that the runoff will be between* $b$ *and* $c$*. Hence* $a$ *has a* 51% *chance of winning,* $b$ *a* 49% *chance of winning, and* $c$ *a* 0% *chance of winning.*

*If* $\mathcal{USR}$ *ranks* $c$ *first with her two votes, then there is a* 98% *chance that the runoff will be between* $b$ *and* $c$ *and a* 2% *chance that the runoff will be between* $a$ *and* $c$*. Hence* $a$ *has a* 0% *chance of winning,* $b$ *a* 98% *chance of winning, and* $c$ *a* 2% *chance of winning.*

*Hence, in any case,* $\mathcal{USR}$ *does not produce a vote in which the first choice is the outcome with the highest chance of winning. Then, for a user in this situation, the strategy of producing a list ordered by probability cannot be incentive compatible because it is impossible to follow. This is true regardless of the system of rewards and punishments applied.*

*Further note that this example also applies to the Smith-IRV system, namely the system of performing IRV but checking before each round whether there is a Condorcet winner and selecting this outcome to be the winner if there is.*

**Proposition 5.** *Suppose we have a monotonic voting system. Then whatever a user's priors* $Vot(\mathcal{USR})$ *about the votes she does not control, there exists a vote* $v_i$ *that she can cast so that the highest ranked choice of* $v_i$ *is the outcome with the highest probability of winning.*

*Proof.* Note that there are a finite number of permutations of the outcomes. Each permutation, if submitted as a vote, yields a percentage of each of its outcomes winning. Take the permutation

$$v_1 : a_{\sigma(1)} > \ldots > a_{\sigma(n)}$$

that maximizes the probability that its first choice wins. A priori, it is possible that some lower ranked choice $a_{\sigma(i)}$ has an even higher probability of winning, $p_{a_{\sigma(i)}} > p_{a_{\sigma(1)}}$. Then consider the situation where the user votes

$$v_2 : a_{\sigma(i)} > a_{\sigma(1)} > \ldots > a_{\sigma(i-1)} > a_{\sigma(i+1)} > \ldots > a_{\sigma(n)}.$$

By monotonicity, for any collection of votes in $Vot(\mathcal{USR})$ in which $a_{\sigma(i)}$ wins if $\mathcal{USR}$ votes $v_1$, $a_{\sigma(i)}$ still wins if $\mathcal{USR}$ votes $v_2$. Hence $v_2$ gives a vote for which the first choice has a probability of winning greater than $p_{a_{\sigma(1)}}$, contradicting the assumed maximality of $a_{\sigma(1)}$. □

**Example 5.** *We see that even with a monotonic voting system, it is still possible to have situations where there is no honest vote. Suppose that $\mathcal{USR}$ controls three votes, and her probability distribution for the other voters is given as described below. Then, note that in each of the Ranked Pairs, Minmax, Kemeny-Young, and Schulze systems, in each scenario the outcome is given as in the tables below. However, these systems are Condorcet and monotonic, (see for example [20] for a discussion of why Ranked Pairs possesses these properties).*

- *Situation 1: There is a* 10% *chance that there will be*
  - 14 *votes for* $a > b > c$
  - 3 *votes for* $a > c > b$
  - 15 *votes for* $b > c > a$
  - 16 *votes for* $c > a > b$

| $\mathcal{USR}$ *vote* | $a > b > c$ | $a > c > b$ | $b > a > c$ | $b > c > a$ | $c > a > b$ | $c > b > a$ |
|---|---|---|---|---|---|---|
| *outcome* | $a$ | $c$ | $a$ | $c$ | $c$ | $c$ |

- *Situation 2: There is a* 32% *chance that there will be*
  - 12 *votes for* $a > b > c$
  - 3 *votes for* $a > c > b$
  - 17 *votes for* $b > c > a$
  - 16 *votes for* $c > a > b$

| $\mathcal{USR}$ *vote* | $a > b > c$ | $a > c > b$ | $b > a > c$ | $b > c > a$ | $c > a > b$ | $c > b > a$ |
|---|---|---|---|---|---|---|
| *outcome* | $c$ | $c$ | $b$ | $b$ | $c$ | $c$ |

- *Situation 3: There is a* 17% *chance that one is in a situation where* $b$ *wins regardless of the vote of* $\mathcal{USR}$*, such as*
  - 48 *votes for* $b > a > c$
- *Situation 4: There is a* 26% *chance that there will be*
  - 13 *votes for* $a > b > c$
  - 6 *votes for* $a > c > b$
  - 18 *votes for* $b > c > a$
  - 11 *votes for* $c > a > b$

| $\mathcal{USR}$ *vote* | $a > b > c$ | $a > c > b$ | $b > a > c$ | $b > c > a$ | $c > a > b$ | $c > b > a$ |
|---|---|---|---|---|---|---|
| *outcome* | $a$ | $a$ | $a$ | $b$ | $c$ | $b$ |

- *Situation 5: There is a* 15% *chance that one is in a situation where* $a$ *wins regardless of the vote of* $\mathcal{USR}$*, such as*
  - 48 *votes for* $a > b > c$

*Then, we summarize which scenarios and which user votes give each outcome:*

| $\mathcal{USR}$ \vote outcome | $a$ | $b$ | $c$ |
|---|---|---|---|
| $a > b > c$ | *1,4,5* | *3* | *2* |
| $a > c > b$ | *4,5* | *3* | *1,2* |
| $b > a > c$ | *1,4,5* | *2,3* | |
| $b > c > a$ | *5* | *2,3,4* | *1* |
| $c > a > b$ | *5* | *3* | *1,2,4* |
| $c > b > a$ | *5* | *3,4* | *1,2* |

*In percent chance of winning:*

| $\mathcal{USR}$ \vote outcome | $a$ | $b$ | $c$ |
|---|---|---|---|
| $a > b > c$ | 51% | 17% | 32% |
| $a > c > b$ | 41% | 17% | 42% |
| $b > a > c$ | 51% | 49% | 0% |
| $b > c > a$ | 15% | 75% | 10% |
| $c > a > b$ | 15% | 17% | 68% |
| $c > b > a$ | 15% | 43% | 42% |

**Remark 2.** *Consider a monotonic voting system and suppose one begins with $a > b > c$ as the vote guaranteed by Proposition 5 that places the highest ranked choice first. Then one could attempt to obtain an honest vote by permuting two options at a time: e.g. if $a > b > c$ yields $p_c > p_b$ then try $a > c > b$. By monotonicity, it is still the case that $p_c > p_b$, but it is possible that $p_c > p_a$. If this is the case one can try $c > a > b$. By monotonicity, it is then the case that $p_c > p_a$, but it is possible that $p_b > p_a$, etc. Continuing like this any prior that $\mathcal{USR}$ might have that results in no possible honest vote in a monotonic voting system (at least with three outcomes) must have a particular form. Then, we see in Example 5 that that form can be realized in the Ranked Pairs, Minmax, Kemeny-Young, and Schulze systems.*

**Question 1.** *Does there exist any Condorcet voting system in which it is always possible to provide an "honest vote", i.e. a ranking in order by the probability that each outcome wins, once your vote is taken into account and for an arbitrary prior on the distribution of the other votes?*

# 9 Conclusion

In this work we have begun to look at a version of social choice theory in the context where participants in an election are motivated by a Schelling game rather than by their own candidate preferences. We have provided a definition for what "honesty" can mean in this framework, and we have observed that many of the subtleties of traditional social choice theory regarding tactical voting have analogs here.

Finally, we examined several possibilities for voting and payout systems in this context and made heuristic arguments on which choices might be appropriate. We argued that Condorcet voting systems are likely well suited to Schelling

games, particularly if voters expect that there will typically be Condorcet winners. (*Which* Condorcet system is appropriate is less clear, though we have made some arguments for Ranked Pairs. Particularly we observed in Section 6.2 that if used with our candidate payoff systems Ranked Pairs makes failed attacks relatively expensive for attackers, and we saw in Proposition 5 that monotonicity is related to the existence of honest votes.) In parallel, we have considered corresponding payoff systems; there we saw that by increasing the weighting of narrowly decided pairs or not, we had a tradeoff between preserving properties related to incentive compatibility versus increasing the expense of an attack.

# References


[1] John Adler, Ryan Berryhill, Andreas G. Veneris, Zissis Poulos, Neil Veira, and Anastasia Kastania. ASTRAEA: A decentralized blockchain oracle. *2018 IEEE Confs on Internet of Things, Green Computing and Communications, Cyber, Physical and Social Computing, Smart Data, Blockchain, Computer and Information Technology, Congress on Cybermatics*, 2018.

[2] Kenneth J. Arrow. *Social Choice and Individual Values*. Yale University Press, 2012.

[3] J. Bartholdi, C. A. Tovey, and M. A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6(2):157–165, 1989.

[4] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia. *Handbook of Computational Social Choice*. Cambridge University Press, New York, NY, USA, 1st edition, 2016.

[5] Markus Brill and Felix Fischer. The price of neutrality for the ranked pairs method. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, AAAI'12, pages 1299–1305. AAAI Press, 2012.

[6] Vitalik Buterin. The p+epsilon attack. `https://blog.ethereum.org/2015/01/28/p-epsilon-attack/`. January 2015.

[7] C.L. Dodgson. A method for taking votes on more than two issues. pages 224–234, 01 1876. reprinted in Duncan Black, The Theory of Committees and Elections (Cambridge: Cambridge University Press, 1963).

[8] Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. On distance rationalizability of some voting rules. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK '09, pages 108–117, New York, NY, USA, 2009. ACM.

[9] Piotr Faliszewski, Edith Hemaspaandra, and Lane A. Hemaspaandra. How hard is bribery in elections? *J. Artif. Int. Res.*, 35(1):485–532, July 2009.

[10] William George and Clément Lesaege. A smart contract oracle for approximating real-world, real number values. To appear in *Proceedings of Tokenomics International Conference on Blockchain Economics, Security, and Protocols*, 2019.

[11] Allan Gibbard. Manipulation of Voting Schemes: A General Result. *Econometrica*, 41(4):587–601, July 1973.

[12] Lane A. Hemaspaandra, Rahman Lavaee, and Curtis Menton. Schulze and ranked-pairs voting are fixed-parameter tractable to bribe, manipulate, and control. *Annals of Mathematics and Artificial Intelligence*, 77(3-4):191–223, August 2016.

[13] Clément Lesaege and Federico Ast. Kleros: Short paper v1.0.5. `https://kleros.io/assets/whitepaper.pdf`. January 2018.

[14] Hervé Moulin. Condorcet's principle implies the no show paradox. *Journal of Economic Theory*, 45(1):53 – 64, 1988.

[15] Mark Allen Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187 – 217, 1975.

[16] T.C. Schelling. *The Strategy of Conflict*. Harvard University Press, 1980.

[17] W.C. Stirling. *Theory of Social Choice on Networks: Preference, Aggregation, and Coordination*. Cambridge University Press, 2016.

[18] Lars-Gunnar Svensson. The proof of the Gibbard-Satterthwaite theorem revisited, 1999.

[19] Paul Sztorc. Truthcoin: Peer-to-peer oracle system and prediction marketplace. `https://www.truthcoin.info/papers/truthcoin-whitepaper.pdf`. Version 1.5, December 2015.

[20] T. N. Tideman. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, 4(3):185–206, Sep 1987.

[21] Vitalik Buterin. SchellingCoin: A minimal-trust universal data feed. Ethereum blog: `https://blog.ethereum.org/2014/03/28/schellingcoin-a-minimal-trust-universal-data-feed/`. March 2014.

[22] Tiance Wang, Paul W. Cuff, and Sanjeev Kulkarni. Condorcet methods are less susceptible to strategic voting. 2013.

[23] Lirong Xia and Vincent Conitzer. Generalized scoring rules and the frequency of coalitional manipulability. pages 109–118, 01 2008.