# Reflections on Vote Manipulation

Jan van Eijck, Floor Sietsma, and Sunil Simon

CWI, Amsterdam, {jve,f.sietsma,s.e.simon}@cwi.nl

The notion of non-manipulability (or: strategy-proofness) used in the famous Gibbard-Satterthwaite theorem is too strong to make useful distinctions between voting rules. We explore alternative definitions and suggest how these can be used to classify voting rules.

If $A$ is a finite set of alternatives, with $|A| > 2$, then an $A$-*ballot* is a linear ordering of $A$. Let $\{1, \ldots, n\}$ be a set of voters. An $(A, n)$-*profile* is an $n$-tuple of $A$-ballots. If $\mathbf{P}$ is an $(A, n)$-profile, then $\mathbf{P}$ can be written as $(\succ_1, \ldots, \succ_n)$. $\succ_i$, the $i$-th component of profile $(\succ_1, \ldots, \succ_n)$, is the ballot of voter $i$. $\succ_i$ expresses "what voter $i$ wants."

If $\mathbf{P}(A)$ is the set of all $(A, n)$-profiles, for given $n \in \mathbb{N}$, then a function $V : \mathbf{P}(A) \to A$ is a *resolute voting rule* for $A$, and a function $V : \mathbf{P}(A) \to \mathcal{P}^+(A)$ is a *voting rule* for $A$.

Let $\mathbf{P} \sim_i \mathbf{P}'$ express that $\mathbf{P}$ and $\mathbf{P}'$ differ only in the ballot of voter $i$. Then the standard definition of non-manipulability is given by:

> A resolute voting rule $V$ is *non-manipulable (NM)* (or: *strategy-proof*) if $\mathbf{P} \sim_i \mathbf{P}'$ implies $V(\mathbf{P}) \succeq_i V(\mathbf{P}')$, where $\succeq_i$ denotes the $i$-preference in $\mathbf{P}$.

A voting rule $V$ is *non-imposed* (NI) if any candidate can be a winner: $\forall a \in A \exists \mathbf{P} : a \in V(\mathbf{P})$. The famous Gibbard/Satterthwaite theorem (GS, [2, 3]) says that *any resolute voting rule that is NM and that is NI is a dictatorship.*

We take the fact that GS has an easy proof as a sign that the notion of manipulability (forbiding *any* adjustment of the vote) is too strong, and we investigate what happens if we distinguish between *reasonable* adjustments and *perverse* adjustments. Following [1], we define the *knights* and *knaves* of a voter $i$, given profile $\mathbf{P}$ and voting rule $V$:

> The *knights* of a voter $i$, given profile $\mathbf{P}$ and resolute voting rule $V$, are the alternatives that are above $V(\mathbf{P})$ on the $i$-ballot. The *knaves* of a voter $i$, given profile $\mathbf{P}$ and resolute voting rule $V$, are the alternatives that are below $V(\mathbf{P})$ on the $i$-ballot.

Thus, if $i$ has ballot $a \succ b \succ c \succ d$ in $\mathbf{P}$, and the outcome of the vote is $c$, then $a$ and $b$ are knights of $i$ in $\mathbf{P}$, and $d$ is a knave of $i$ in $\mathbf{P}$.

Using this, we can distinguish between *benign adjustments* and *perversions* by means of a check whether the manipulation involves knight demotion or knave promotion.

A resolute voting rule $V$ is *demotion pervertible* (DP) if there exists an $i$-minimal pair of profiles $\mathbf{P}, \mathbf{P}'$ such that (i) $V(\mathbf{P}) \prec_i V(\mathbf{P}')$, and (ii) $\exists x : V(\mathbf{P}) \prec_i$

$x \prec_i' V(\mathbf{P})$. A resolute voting rule $V$ is *NDP* (non-demotion-pervertible) if $V$ is not DP.

Note that the demotion of knight $x$ from above $V(\mathbf{P})$ to a new position below $V(\mathbf{P})$ is the perversion. For example, suppose $i$ has ballot *abcd* in $\mathbf{P}$, and the outcome of the vote is $c$, and $\mathbf{P} \sim_i \mathbf{P}'$ where $i$ has ballot *bcad* in $\mathbf{P}'$, and the outcome of the vote in $\mathbf{P}'$ is $b$. Then $V$ is demotion pervertible, for we have that $V(\mathbf{P}) = c \prec_i b = V(\mathbf{P}')$, and $V(\mathbf{P}) = c \prec_i a \prec_i' c = V(\mathbf{P})$, that is to say, $a$ was demoted from a knight to a knave position (from the perspective of $\mathbf{P}$).

In a similar way, we can define (non)-promotion pervertibility. A promotion perversion would be the promotion of a knave from below $V(\mathbf{P})$ to a new position above $V(\mathbf{P})$.

An $i$-minimal pair of profiles $\mathbf{P}$, $\mathbf{P}'$ *invites decency towards knights* if the following hold: (i) $V(\mathbf{P}) \prec_i V(\mathbf{P}')$ implies $\forall x : V(\mathbf{P}) \prec_i x \Rightarrow V(\mathbf{P}) \preceq_i' x$, and vice versa: (ii) $V(\mathbf{P}) \succ_i' V(\mathbf{P}')$ implies $\forall x : V(\mathbf{P}') \prec_i' x \Rightarrow V(\mathbf{P}') \preceq_i x$.

This says: "If the shift from $\succeq_i$ to $\succeq_i'$ is an improvement, then no knight was demoted", and similarly in the other direction. In a similar fashion, we can define decency towards knaves ("don't promote them to where they do not fit").

The above definition of NDP is not suitable yet to classify voting rules, for what matters is not whether the vote can be perverted by demotion, but whether such perversion is essential. The following definition takes this into account.

> A voting rule $V$ is not single-winner demotion pervertible if for any two profiles $\mathbf{P} \sim_i \mathbf{P}'$, where $V(\mathbf{P}) \prec_i V(\mathbf{P}')$ and there is an $x \succ_i V(\mathbf{P}) \succ_i' x$, there exists another profile $\mathbf{Q}$ such that $\mathbf{P} \sim_i \mathbf{Q}$ and $V(\mathbf{Q}) = V(\mathbf{P}')$.

What this says is that perversions of the vote may be possible, but they are never essential to achieve the goal: for every vote perversion there is a non-perverted alternative that works just as well.

**Theorem 1.** *The plurality rule is not single-winner promotion or demotion pervertible.*

*Proof.* Suppose $\mathbf{P} \sim_i \mathbf{P}'$ and $V(\mathbf{P}) \prec_i V(\mathbf{P}')$. Since $V(\mathbf{P}) \neq V(\mathbf{P}')$, either $i$'s top candidate in $\mathbf{P}$ is $V(\mathbf{P})$ or $i$'s top candidate in $\mathbf{P}'$ is $V(\mathbf{P}')$. In the first case, $V(\mathbf{P}) \succeq_i V(\mathbf{P}')$, which contradicts our assumption. In the second case, let $\mathbf{Q}$ be the profile with $i$'s ballot identical to that in $\mathbf{P}$, with the only difference being that $V(\mathbf{P}')$ got moved to the top. Since $V(\mathbf{P}') \succ_i V(\mathbf{P})$, this cannot involve promotion or demotion and since in the plurality rule only the top element of the ballot counts, $V(\mathbf{Q}) = V(\mathbf{P}')$. □

The following theorem can be proved by example:

**Theorem 2.** *The weak Condorcet rule is single-winner promotion pervertible.*

What these examples show is that our refined notions of manipulability can be used to make useful distinctions. In future work we intend to further classify positional voting rules with respect to the new notions of manipulability.

## References

1. Jan van Eijck. A geometric look at manipulation. In J. et al. Leite, editor, *CLIMA XII 2011*, volume 6814 of *Lecture Notes in Artificial Intelligence*, pages 92–104. Springer, 2011.
2. A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41:587–601, 1973.
3. M.A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.