3

Resolute Voting Rules

3.1 The Gibbard–Satterthwaite Theorem

For the sake of keeping this chapter relatively self-contained, we begin by restating the definition of manipulability in our present context of resolute voting rules for three or more alternatives. As we are considering only resolute procedures, there is no need to use the phrase "single-winner" when speaking of manipulability.

Definition 3.1.1. In the context of linear or non-linear ballots, a resolute voting rule V is *manipulable* if there exists a profile $\mathbf{P} = (R_1, \ldots, R_n)$, which we think of as giving the true preferences of the n voters, and another ballot Q_i, which we think of as a disingenuous ballot from voter i such that, letting $\mathbf{P}' = (R_1, \ldots, R_{i-1}, Q_i, R_{i+1}, \ldots, R_n)$, we have:

$$V(\mathbf{P}')P_i V(\mathbf{P})$$

If V is not manipulable, then V is *non-manipulable.*

A natural question at this stage is the following: For a given set of alternatives and a given set of voters, exactly which resolute voting rules are non-manipulable? A moment's thought suggests two special cases in which non-manipulable, resolute voting rules certainly exist:

(1) If there is only one voter, then we can take the alternative at the top of his or her ballot as the winner.
(2) If there are only two alternatives and an odd number of voters, then we can use majority rule.

In fact, both ideas apply in the general case where there are several alternatives and several voters. That is, we could implement idea (1) by choosing a particular voter and ignoring all other voters except this one. Or we could implement idea (2) by choosing a particular pair of alternatives and ignoring all other alternatives except these two.

In the context of linear ballots, *all* non-manipulable resolute voting rules arise as natural generalizations of these two observations. However, we put off this discussion until we present some characterization theorems in Chapter 7, and turn instead to the main result of this chapter (indeed, the main result in the study of manipulability of voting systems).

In terms of history, it was in the early 1970s that the philosopher Allan Gibbard and the economist Mark Satterthwaite independently established the theorem bearing their names. Gibbard's result appeared in his article "Manipulation of Voting Schemes: a General Result," *Econometrica 41* (1973) 587–601; Satterthwaite included the result in his 1973 Ph.D. thesis, written at the University of Wisconsin, and he published it later in an article entitled "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory* 10 (1975), 187–217.

The version of the Gibbard–Satterthwaite theorem we state here assumes Pareto, but we later give the easy argument showing that the assumption of non-imposition suffices. Both Gibbard and Satterthwaite handled the case where ties in the ballots are allowed, as we do in Section 3.

Theorem 3.1.2 (The Gibbard–Satterthwaite Theorem for Linear Ballots). *In the context of linear ballots, if* n *is a positive integer and* A *is a set of three or more alternatives, then any resolute voting rule for* (A, n) *that is non-manipulable and that satisfies Pareto is a dictatorship.*

The more important a theorem is, the harder we should look for explanations of why it is true. And the Gibbard–Satterthwaite theorem is extremely important. Thus we find, as we should, a number of different proofs in the literature over the past three decades. Examples include Gärdenfors (1977), Schmeidler and Sonnenschein (1978), Feldman (1979c), Barberá (1983), Benoit (2000), Arunava Sen (2001), and Taylor (2002). For other book-length treatments of manipulability, see Moulin (1983 and 1985) and Riker (1986).

Before beginning the proof of Theorem 3.1.2, let's set the stage by talking about the overall strategy, which is, in fact, the same strategy underlying most proofs of Arrow's impossibility theorem. Assume, then, that V satisfies Pareto and is non-manipulable. Our goal is to "find" the voter i who is, in fact, the dictator for V.

The first part of the strategy is to see if we can restrict ourselves to talking about *pairs* of alternatives. For example, the definition of a dictatorship speaks about the whole set A of alternatives and the effect of a dictator arranging all of these in some order on his or her ballot. But what can we say if we only know that for two particular alternatives – call them a and b – the dictator ranks a over b? We certainly cannot conclude that a is the winner. But we *can* conclude that b is a non-winner. Moreover, if a voter had this power for every pair of alternatives, then he or she would, of necessity, be a dictator.

Now, let's consider the other extreme. Instead of looking at a single voter (a dictator, say), let's look at the whole set N of voters. Suppose they all rank a over b on their ballots. What can we conclude? Again, we certainly cannot conclude that a is the winner. But, if Pareto holds, then we can conclude that b is a non-winner.

This suggests the following. If X is a set of voters, call X a "dictating set" if it has the following property: For every profile **P** and every pair a, b of alternatives, if everyone in X ranks a over b, then $V(\mathbf{P}) \neq b$.

Our overall task is now set. We are beginning with the knowledge that the set N itself is a dictating set – this is precisely what Pareto asserts – and we want to find a single voter i such that $\{i\}$ is a dictating set, because this is equivalent to asserting that voter i is a dictator. One way to accomplish this, as we show in the following paragraph, is to prove a lemma that guarantees that if a dictating set is split into two pieces, then one of the two pieces is again a dictating set.

Such a lemma will immediately yield the desired result, because we know that N is a dictating set, and so one exists, and thus there must be one of minimal size, which must then be a singleton by the proposed lemma. Moreover, if the result we are trying to prove is true, then the lemma must also be true because the presence of a dictator ensures that the dictating sets are precisely the sets to which he or she belongs. Hence, our attempts to prove the lemma are not doomed before we start unless the theorem itself is false.

So how do we prove such a lemma? The key is in obtaining one more refinement. Being a dictating set means that for every pair of alternatives, a certain thing happens: Namely, if everyone ranks one of the alternatives over another, the latter is not the winner. It certainly makes sense to ask of a *given pair* a, b of alternatives if this same thing happens. Formalizing this yields the following.

Definition 3.1.3. If X is a set of voters, and a and b are two distinct alternatives in the set A, then *X can use* a *to block* b, denoted aXb, if, for every profile **P** in which all the voters in X rank a over b on their ballots, $V(\mathbf{P}) \neq b$. The set X is a *dictating set* if aXb for every distinct pair a, b of alternatives in A.

Definition 3.1.3 takes place in the context of a fixed resolute voting rule V, and our notation "aXb" and terminology "dictating set" could have been chosen to reflect that dependence on V. For example, we could have chosen to write

"aXb (mod V)" and to speak of a "dictating set for V." The approach in Definition 3.1.3, however, should cause no confusion.

Up to this point, we could as easily have been talking about proving Arrow's impossibility theorem as the Gibbard–Satterthwaite theorem. To bring manipulability into the argument, we could go after the desired lemma directly or – and this is the route we will pursue – we could identify an election-theoretic consequence of manipulability that is combinatorially easier to apply. Consider the following.

Definition 3.1.4. A resolute voting rule V satisfies *down-monotonicity* provided that, for every profile **P**, if **P**′ is the profile obtained from **P** by having one voter move one losing alternative down one spot on his or her ballot, then V(**P**′) = V(**P**).

Notice that if V satisfies down-monotonicity and V(**P**) = x, then V(**P**′) = x whenever **P**′ is derived from **P** by having *several* voters move *several* losing alternatives down *several* slots on their ballots. This observation will be important in applying down-monotonicity. One can also show that down-monotonicity implies monotonicity (see Exercise 2).

In our proof of the Gibbard–Satterthwaite theorem, the only direct appeal that we will make to manipulability is in the following lemma.

Lemma 3.1.5. *Every resolute voting rule that is non-manipulable satisfies down-monotonicity.*

Proof: Assume that down-monotonicity fails for the resolute voting rule V. Then there exist two profiles **P** and **P**′ and an alternative y such that:

(1) In **P**, voter i ranks y directly over x, V(**P**) = w, and $w \neq y$ (that is, y is the losing alternative that voter i will be moving down).
(2) **P**′ differs from **P** only in that voter i has interchanged the position of x and y on his or her ballot and yet V(**P**′) = v for some $v \neq w$.

The situations described in (1) and (2) are pictured below:

P		**P**′	
ballot i	winner	ballot i	winner
y		x	
x	$w \neq y$	y	$v \neq w$

Assuming (1) and (2), we will show that the system can, in fact, be manipulated.

Case 1: v is over w on voter i's ballot in **P**.

In this case, we can regard voter i's ballot in **P** as giving his or her true preferences. Thus, if he or she submits the sincere ballot **P**, w is the winner, although he or she prefers v to w. But if he or she submits a disingenuous ballot (the one in **P**′), then v is, in fact, the winner, and he or she prefers v to w according to his or her true preferences given in **P**.

Case 2: w is over v on voter i's ballot in **P**′.

In this case, we can regard voter i's ballot in **P**′ as giving his or her true preferences. Thus, if he or she submits the sincere ballot **P**′, v is the winner, although he or she prefers w to v. But if he or she submits a disingenuous ballot (the one in **P**), then w is, in fact, the winner, and he or she prefers w to v, according to his or her true preferences given in **P**′.

Case 3: Otherwise.

In this case, w is over v on voter i's ballot in **P**, and v is over w on voter i's ballot **P**′. But this means that $w = y$ and $v = x$ contracting our assumption that $y \neq w$. □

To motivate the next lemma, suppose that we have a set X of voters and two distinct alternatives a and b. Let's think about how we would verify that aXb. If we appeal directly to Definition 3.1.3, then we have to examine *every* profile in which all the voters in X rank a over b, and check to see that b is, in fact, a non-winner in each of these elections.

It would be nice if, in the definition aXb, we could replace the initial universal quantifier ("for every profile **P**....") with an existential one ("there exists a profile **P**...."). This is, in fact, possible, if we also make two other changes simultaneously: We must require that everyone not in X place b over a, and instead of asserting that V(**P**) $\neq b$, we must get the stronger conclusion that V(**P**) $= a$ (which implies that V(**P**) $\neq b$ because we are assuming resoluteness). These comments are formalized in Lemma 3.1.6 ("the existence lemma") below.

For the remaining five lemmas, we assume that V is a resolute voting rule for (A, n) in the context of linear ballots, and that V satisfies down-monotonicity and Pareto.

Lemma 3.1.6 (The Existence Lemma). *Suppose that X is a set of voters and that* a *and* b *are two distinct alternatives in A. Then, in order to show that* aXb, *it suffices to produce one profile* **P** *for which:*

(1) everyone in X ranks a *over* b,
(2) everyone else ranks b *over* a, *and*
(3) $V(\mathbf{P}) = $ a.

Proof: Suppose we have such a profile **P** but aXb fails. Then we also have a profile **P**′ in which everyone in X ranks a over b, and $V(\mathbf{P}') = b$. With **P**′, some voters not in X might also rank a over b, but because we are assuming down-monotonicity, we can obtain a new profile **P**″ by having every such voter move the losing alternative a down below b, and still have $V(\mathbf{P}'') = b$. Thus, with both **P** and **P**″, (1) and (2) above hold, with $V(\mathbf{P}) = a$ and $V(\mathbf{P}'') = b$.

Now, choose an alternative c that is distinct from a and b and have every voter move c to the bottom of his or her ballot in both profiles. By down-monotonicity, the winner is still a in the first election and still b in the second. Now choose an alternative d that is distinct from a, b, and c (if there is one) and do the same thing. Continuing this, we eventually get two elections having identical sequences of ballots, with alternative a winning the first and alternative b winning the second. This contradiction completes the proof. □

The next lemma is difficult to motivate directly, but it is precisely the statement whose proof is provided by using the Concorcet voting paradox ballots in our present context.

Lemma 3.1.7 (The Splitting Lemma). *Suppose X is a set of voters and that* a, b, *and* c *are distinct alternatives in A. Assume also that* aXb *and that X is partitioned into disjoint subsets Y and Z (one of which may be empty). Then either* aYc *or* cZb.

Proof: Consider any profile **P** in which every voter in Y has a first, b second, and c third; every voter in Z has c first, a second, and b third; and everyone else (i.e., those voters in N–X) has b first, c second, and a third (with all other alternatives below these). We can picture these ballots as follows:

P

ballots of voters in Y	ballots of voters in Z	ballots of voters in N–X
a	c	b
b	a	c
c	b	a
.	.	.
.	.	.
.	.	.

By Pareto, $V(\mathbf{P}) \in \{a, b, c\}$ (everyone, for example, prefers a to d, so $V(\mathbf{P}) \neq d$). Moreover, $V(\mathbf{P}) \neq b$ because aXb by assumption, and everyone in $X = Y \cup Z$ ranks a over b. But now, Lemma 3.1.6 (the existence lemma) shows that if $V(\mathbf{P}) = a$, then aYc, and if $V(\mathbf{P}) = c$, then cZb. □

Lemma 3.1.8. *Suppose X is a set of voters and that* a, b, *and* c *are three distinct alternatives in A. Then*

(1) if aXb, *then* aXc, *and*
(2) if aXb, *then* cXb.

Proof: Notice that in Lemma 3.1.7, we allowed Y or Z to be the empty set. Because Pareto holds, we never have $a\emptyset b$. Conclusions (1) and (2) now follow immediately from Lemma 3.1.7 by letting Y = N and Z = ∅, and then Y = ∅ and Z = N. □

Lemma 3.1.9. *Suppose X is a set of voters and that* aXb *holds for some* a *and* b. *Then X is a dictating set.*

Proof: Assume x and y are distinct alternatives. We show that $x\mathrm{X}y$ must hold.

Case 1. $y \neq a$
Because $a\mathrm{X}b$ and $y \neq a$, we know by Lemma 3.1.8 (1) that $a\mathrm{X}y$. Because $x \neq y$, we can now apply Lemma 3.1.8 (2) to get $x\mathrm{X}y$, as desired.
Case 2. $x \neq b$
Because $a\mathrm{X}b$ and $x \neq b$ we know by Lemma 3.1.8 (2) that $x\mathrm{X}b$. Because $y \neq x$, we can now apply Lemma 3.1.8 (1) to get $x\mathrm{X}y$, as desired.
Case 3. $y = a$ and $x = b$
Because A has three or more elements, we can choose c distinct from a and b. Now, because $a\mathrm{X}b$, we know by Lemma 3.1.8 (1) that $a\mathrm{X}c$, and by Lemma 3.1.8 (2) that $b\mathrm{X}c$. A final application of Lemma 3.1.8 (1) shows that $b\mathrm{X}a$, and so $x\mathrm{X}y$, as desired. □

Lemma 3.1.9 allows us to conclude that a set X of voters is a dictating set if $a\mathrm{X}b$ for even *one* pair of alternatives a and b. With this observation and our discussion preceding Definition 3.1.3 (formalized as Lemma 3.1.11 below), we can conclude the proof of Theorem 3.1.2 with the following two lemmas.

Lemma 3.1.10. *If X is a dictating set and X is split into disjoint subsets Y and Z, then either Y is a dictating set or Z is a dictating set.*

Proof: If a, b, and c are any three distinct alternatives, then we know $a\mathrm{X}b$ because X is a dictating set. But now, by Lemma 3.1.7, either $a\mathrm{Y}c$ or $c\mathrm{Z}b$. By Lemma 3.1.9, Y is a dictating set in the former case, and Z is a dictating set in the latter. □

Lemma 3.1.11. *If X is a dictating set, then there exists a voter* i $\in$ *X such that* {i} *is a dictating set. In particular, because N is a dictating set (by Pareto), there exists a voter who is a dictator for V.*

Proof: This is immediate from Lemma 3.1.10 and the discussion preceding Definition 3.1.3. □

These lemmas complete the proof of the Gibbard–Satterthwaite theorem in the context of linear ballots. In the next section, we handle the case of non-linear ballots, but before moving on, we state (as corollaries) two equivalent versions of the Gibbard–Satterthwaite theorem. The first replaces Pareto with the weaker assumption of non-imposition (and recasts the result to make it a characterization theorem). The second illustrates what the Gibbard–Satterthwaite theorem yields in the context of non-resolute voting rules.

Corollary 3.1.12. *In the context of linear ballots and three or more alternatives, a voting rule is non-manipulable, non-imposed, and resolute iff it is a dictatorship.*

Proof: A dictatorship is clearly non-manipulable, non-imposed, and resolute. For the converse, it suffices to show that if V is non-manipulable, non-imposed, and resolute, then V satisfies Pareto.

Suppose, for contradiction, that Pareto fails and let **P** be a profile in which every voter ranks a over b, but V(**P**) = b. Because V is non-manipulable, it satisfies down-monotonicity, so we lose no generality in assuming that every voter has plunged every alternative other than a and b below b. Hence, every voter has a at the top of his or her ballot.

Because V is non-imposed, we can choose a profile **P**′ with V(**P**′) = a. Now, one by one, replace the ballots in **P** with those of **P**′ (so voter 1's ballot in **P** is replaced by his or her ballot in **P**′, then voter 2's ballot in **P** is replaced by his or her ballot in **P**′, and so on). At some point, the election outcome becomes a for the first time. If that last change of ballot is voter i, then we can assume his or her original ballot – with a on top – represents his or her true preferences. But then his or her change to a disingenuous ballot changes the result from something other than his or her top choice to his or her top choice a. This is an instance of manipulation, and completes the proof. □

Corollary 3.1.13. *In the context of linear ballots and three or more alternatives, every non-imposed (not necessarily resolute) voting rule for* (*A*, n) *that is not a dictatorship is manipulable in the sense there exist profiles* **P** *and* **P**′ *and a* *voter* i *such that* $\mathbf{P}|N - \{\mathrm{i}\} = \mathbf{P}'|N - \{\mathrm{i}\}$ *and voter* i, *whose true preferences we take to be given by his ballot in* $\mathbf{P}$, *prefers the election outcome X from* $\mathbf{P}'$ *to the election outcome Y from* $\mathbf{P}$ *in the following sense:*

$$max_i(X - Y, \mathbf{P})P_i\, min_i(Y, \mathbf{P}) \text{ or } max_i(X, \mathbf{P})P_i\, min_i(Y - X, \mathbf{P}).$$

Proof: Assume that V is a non-imposed (not necessarily resolute) voting rule that is not a dictatorship. Fix a linear ordering L of the set A of alternatives, and let V′ be the resolute voting rule in which V′(**P**) is the (unique) L-largest element of V(**P**). Clearly, V′ is non-imposed and not a dictatorship, and so V′ is manipulable in the Gibbard–Satterthwaite sense. Thus, there exists an election in which some voter who has x over y on his or her ballot can unilaterally change the outcome from y to x by submitting a disingenuous ballot. But then, with V, the set Y of winners with his or her sincere ballot included y, and the set X of winners with his disingenuous ballot included x. Moreover, we can't have both x and y in X ∩ Y. If $x \notin$ Y, then

$$\max_i(\mathrm{X} - \mathrm{Y}, \mathbf{P})\, \mathrm{R}_i\, x\, \mathrm{P}_i\, y\, \mathrm{R}_i \min_i(\mathrm{Y}, \mathbf{P}).$$

If $y \notin$ X, then

$$\max_i(\mathrm{X}, \mathbf{P})\, \mathrm{R}_i\, x\, \mathrm{P}_i\, y\, \mathrm{R}_i \min_i(\mathrm{Y} - \mathrm{X}, \mathbf{P}).$$

□

3.2 Ties in the Ballots

Both Gibbard's and Satterthwaite's original proofs handled the case where ties in the ballots are allowed. Later, it was realized that one could first prove the result for linear ballots, and then derive the general case from this, as we now do. First, we need a definition in the context of non-linear ballots.

Definition 3.2.1. In the context of non-linear ballots, a voter is a *weak dictator* for a resolute voting rule if the unique winner of every election is one of the alternatives that he or she has tied for top on his or her ballot. V is a *weak dictatorship* if there exists a voter who is a weak dictator for V.

Theorem 3.2.2 (The Gibbard–Satterthwaite Theorem for Non-Linear Ballots). *In the context of non-linear ballots, if* n *is a positive integer and A is a set of three or more alternatives, then any resolute voting rule for* (A, n) *that is non-imposed and non-manipulable is a weak dictatorship.*

Proof: We derive Theorem 3.2.2 from Corollary 3.1.12 as follows. Let V′ be the restriction of V to linear ballots. Then clearly V′ is non-manipulable because

V is. We also claim that V′ satisfies unanimity, as we now show. So suppose for contradiction that **P** is a profile in which every voter has a linear ballot with the same alternative x at the top and that $\mathrm{V}(\mathbf{P}) \neq x$. Choose a profile $\mathbf{P}'$, with ties, perhaps, such that $\mathrm{V}(\mathbf{P}') = x$. Now, starting with the profile **P** of linear ballots, change them one by one into $\mathbf{P}'$ until x becomes the winner. This last change of ballot (by voter i, say) represents a manipulation of V because voter i, whose true preferences, we assume, are given by his or her linear ballot in **P** with x at the top, changed ballots and secured a win for his or her top choice x.

It now follows from Corollary 3.1.12 that V′ is a dictatorship. Thus, we can assume that if the ballots are linear, then voter i's top choice is the unique winner. We claim that voter i is a weak dictator for V.

Suppose voter i is not a weak dictator for V. Then there exists a profile **P** for which $\mathrm{V}(\mathbf{P}) = x$, but x is not among the alternatives that voter i has tied on top of his or her ballot. One by one, move x up on every other ballot in **P** so that x is alone at the top. Then x remains the winner, or undoing such a move would represent a successful manipulation by that voter. Similarly, we can one by one break all the ties in these other ballots and still have x the winner, lest undoing such a change would yield a voter his or her top choice x, thus again constituting manipulation. Finally, we can break all the ties in voter i's ballot, in which case some alternative that was in his or her top block becomes the winner. Thus, if the original ballot represented voter i's true preferences, then he or she has gained by the disingenuous breaking of ties in the ballot. This completes the proof. □

The converse of Theorem 3.2.2 fails in the following sense: It is not true, in the context of non-linear ballots, that every weak dictatorship is non-manipulable; see Exercise 3. However, some weak dictatorships are non-manipulable in the context of non-linear ballots; see Exercise 4.

3.3 The Equivalence of Arrow's Theorem and the Gibbard–Satterthwaite Theorem

In geometry, we say that two versions of the parallel postulate are equivalent if each becomes a theorem when the other is added as an axiom to Euclid's original four. Similarly, we say that two versions of the axiom of choice are equivalent if each becomes a theorem when the other is added as an axiom to the Zermello–Frankel axioms for set theory.

The reasons these assertions have formal content is that the results whose equivalence is being claimed are independent of the remaining axioms (assuming the consistency of the remaining axioms). Absent this condition of independence, the theorem asserting that $2+2=4$ would qualify as being equivalent to Andrew Wiles' elliptic curve result that settled Fermat's last theorem (each being provable from the standard axioms of set theory with the other added – or not added, as it turns out).

Equivalence, however, is also used in an informal sense inspired by the formal notion above. We say that two theorems are equivalent if each is "easily derivable" from the other, where the ease of the derivation is measured (intuitively) relative to the difficulty of the stand-alone proofs of the theorems whose equivalence is being asserted. It is in this informal sense that we want to ask about the equivalence of Arrow's theorem and the Gibbard–Satterthwaite theorem.

Initially, we work in the context of linear ballots, and we begin with the original version of Arrow's theorem that took place in the context of social welfare functions (or social choice functions that satisfy transitive rationality, but these naturally correspond to social welfare functions). In what ways do the two theorems differ?

(1) Arrow's theorem pertains to social welfare functions; the Gibbard–Satterthwaite theorem pertains to voting rules.
(2) The Gibbard–Satterthwaite theorem requires resoluteness as an assumption; Arrow's theorem does not.
(3) Arrow's theorem requires Pareto as an assumption (and fails in the presence of non-imposition as shown by an antidictatorship); the Gibbard–Satterthwaite theorem requires only non-imposition.[15]
(4) Non-manipulability implies monotonicity: IIA does not (as shown again by an antidictatorship).

Issues 1 and 2 have been addressed in the exercises at the end of Chapter 1. We showed there that the voting rule version of Arrow's theorem is equivalent to the social welfare version of Arrow's theorem (and this is a very good model for the use of the word "equivalent"). Moreover, we outlined there a direct proof that a voting rule satisfying Pareto and IIA is resolute.

Issues 3 and 4 seem more delicate, and we deal with them by slightly weakening the statement of each theorem as follows:

(a) We replace "non-imposition" by "Pareto" in the Gibbard–Satterthwaite theorem.

[15] There are versions of Arrow's theorem, dating back to Wilson (1972), that do not assume Pareto. For an explicit treatment where Pareto is replaced by non-imposition, see Saari (1995, p. 87).

(b) We replace "IIA" by "MIIA" in Arrow's theorem, where MIIA is a version of independence of irrelevant alternatives that builds in monotonicity for resolute procedures.

More precisely, a resolute voting rule satisfies MIIA if, whenever $\mathbf{P}$ and $\mathbf{P}'$ are profiles and a and b are distinct alternatives such that $\mathrm{V}(\mathbf{P}) = a$ and $\mathrm{V}(\mathbf{P}') = b$, there is some i such that $a\mathrm{P}_i b$ and $b\mathrm{P}'_i a$. With this, we can now prove the following.

Theorem 3.3.1. *For every set A of three or more alternatives and every* $n \geq 1$, *a voting rule for* (A, n) *that satisfies Pareto is resolute and non-manipulable iff it satisfies MIIA.*

Proof: The conditions on either side of the "iff" hold precisely when the procedure is a dictatorship. But this observation misses the point of what we are trying to establish, and so we proceed with direct arguments of the two implications.

Assume first that V is resolute, non-manipulable, and satisfies Pareto. Suppose that $\mathbf{P}$ and $\mathbf{P}'$ are profiles and a and b are distinct alternatives such that $\mathrm{V}(\mathbf{P}) = a$ and $\mathrm{V}(\mathbf{P}') = b$. We want to show that there is some i such that $a\mathrm{P}_i b$ and $b\mathrm{P}'_i a$. Because V is non-manipulable and resolute, we know that it satisfies down-monotonicity by Lemma 3.1.5 (which was short and completely self-contained).

Thus, for each voter j with $b\mathrm{P}_j a$ and $a\mathrm{P}'_j b$, we can plunge b to the bottom of that ballot in P. Moreover, if the alternatives other than a and b are $c_1, \ldots,$ c_k, we can now, for each ballot in $\mathbf{P}$ and in $\mathbf{P}'$, plunge these other alternatives to the bottom of the ballots in the order c_1, then c_2, etc.

Retaining the names $\mathbf{P}$ and $\mathbf{P}'$ for these now-altered profiles, we have that a and b occupy the top two spots on all ballots in both profiles, all other alternatives occur in the order $c_1, \ldots, c_k$ on every ballot, any ballot that has b over a in $\mathbf{P}$ has b over a in $\mathbf{P}'$ (lest b would have been plunged in $\mathbf{P}$), and $\mathrm{V}(\mathbf{P}) = a$ and $\mathrm{V}(\mathbf{P}') = b$. Because $\mathbf{P} \neq \mathbf{P}'$, we can choose i such that $\mathrm{P}_i \neq \mathrm{P}'_i$. But this means $a\mathrm{P}_i b$ and $b\mathrm{P}'_i a$, as desired.

Now assume that V satisfies MIIA and Pareto. By Exercise 12 in Chapter 1 (a short and self-contained exercise), we know that V is resolute, so it suffices to show that V is non-manipulable. So suppose that $\mathbf{P}$ and $\mathbf{P}'$ are profiles, i is a voter, $\mathbf{P}|\mathrm{N} - \{i\} = \mathbf{P}'|\mathrm{N} - \{i\}$, $\mathrm{V}(\mathbf{P}) = a$ and $b\mathrm{P}_i a$. Then MIIA implies that $\mathrm{V}(\mathbf{P}') \neq b$ because every voter who has a over b in $\mathbf{P}$ still has a over b in $\mathbf{P}'$. Hence, voter i's attempt at manipulation failed. □

If Theorem 3.3.1 convinces the reader that Arrow's theorem is equivalent to the Gibbard–Satterthwaite theorem, then it's unlikely that other results, old or new, will change the reader's mind. On the other hand, if the reader is left unconvinced of the equivalence, then he or she should not conclude that they fail to be equivalent, but simply that our Theorem 3.3.1 was inadequate for the task at hand. The efforts of others (for example, Bernard Monjardet's discussion in Monjardet, 1999), might be more convincing.

We offer two additional remarks here. First, both Arrow and Gibbard and Satterthwaite proved their theorems directly in the context of ballots that allowed ties. Later, as we said, it was realized that the Gibbard–Satterthwaite theorem in this context could easily be derived from the linear-ballot context (as we did in Section 3.2). It turns out that the same is true of Arrow's theorem – see Exercises 5 and 6. Second, in the case of infinitely many voters, the equivalence of Arrow's theorem and the Gibbard–Satterthwaite theorem evaporates unless one replaces individual manipulability with coalitional manipulability (Section 6.2) and, even there, differences arise if there are infinitely many alternatives as well as infinitely many voters (Section 6.4).

3.4 Reflections on the Proof of the Gibbard–Satterthwaite Theorem

A couple of the pieces of the proof of the Gibbard–Satterthwaite theorem are worth isolating. For example, Lemma 3.1.9 actually holds for every binary relation in the following sense (see Blau, 1972, Monjardet, 1978, and Makinson, 1996).

Proposition 3.4.1 (The All-or-None Lemma). *Suppose that A has three or more elements and that ß is an irreflexive binary relation on A that satisfies the following: for any three distinct alternatives $a, b, c \in A$,*

(1) if aßb, *then* aßc, *and*
(2) if aßb, *then* cßb.

Then either $ß = \emptyset$ *or* $ß = A \times A - \Delta$, *where* $\Delta = \{(a, a)\colon a \in A\}$.

Proof: Assume that $ß \neq \emptyset$, and choose $a, b \in A$ such that $aßb$. Now let $(x, y) \in A \times A - \Delta$. We show that $(x, y) \in ß$.

Case 1. $y \neq a$
Because $aßb$, we know by (1) that $aßy$. Because $x \neq y$, we can now apply (2) to get $xßy$ and so $(x, y) \in ß$.
Case 2. $x \neq b$
Because $aßb$, we know by (2) that $xßb$. Because $y \neq x$, we can now apply (1) to get $xßy$ and so $(x, y) \in ß$ again.
Case 3. $y = a$ and $x = b$

Because A has three or more elements, we can choose c distinct from a and b. Now, because $aßb$, we know by (1) that $aßc$, and by (2) that $bßc$. A final application of (1) shows that $bßa$, and so $xßy$. Thus, $(x, y) \in ß$, as desired. This completes the proof. □

The second piece of the proof of the Gibbard–Satterthwaite theorem that is worth isolating for future reference is somewhat more voting theoretic, although there is no voting rule explicitly involved in the following proposition.

Proposition 3.4.2. *Suppose that for each set $X \subseteq N$ we have an antireflexive binary relation, also denoted by X, on the set A of three or more alternatives. Say that a set $X \subseteq N$ is a dictating set if* aXb *for every distinct pair* a *and* b *of alternatives in A, and assume that the set N itself is a dictating set and that* a∅b *fails for every* a *and* b. *Suppose the following holds:*

The splitting condition: If aXb, *and* c *is distinct from* a *and* b, *then for every partition of X into disjoint subsets Y and Z (one of which may be empty), either* aYc *or* cZb.

Then for every dictating set X there exists an $i \in X$ *such that* {i} *is also a dictating set.*

Proof: We first claim that if aXb for some pair a and b of alternatives, then X is a dictating set. To see this, we respectively let $Y = \emptyset$ and then $Z = \emptyset$ in the splitting condition. Because we never have $a\emptyset b$ for any a and b, the assumptions in Proposition 3.4.1 are satisfied by the binary relation X. Thus, either aXb fails for every a and b, which we are assuming is not the case, or X is a dictating set, as desired.

Now, if X is a dictating set, then we can choose $Y \subseteq X$ to be a non-empty dictating set of minimal size among all subsets of X. If Y were not a singleton, then we could once again appeal to the splitting condition to arrive at a proper subset of Y that is, by the first paragraph of this proof, again a dictating set, thus contradicting the minimality of Y. □

In the proof of the Gibbard–Satterthwaite theorem, the relation X was defined in terms of a resolute voting rule V by asserting that aXb holds provided that, for every profile **P** in which every voter in X ranks a over b on his or her ballot, we have $V(\mathbf{P}) \neq b$. But this is only one such possibility. And, although we later make use of Proposition 3.4.2 in the context of non-manipulability, let us here illustrate its use with two different interpretations of "aXb" in the context of Arrow's theorem. In each case, we will assume a version of Pareto and a version of IIA, and then we will prove an "existence lemma" and a "splitting lemma." Proposition 3.4.2 will then immediately give as some kind of dictatorial behavior.[16]

For simplicity, we assume in the next three theorems that we have three or more alternatives, linear ballots, and that the aggregation procedures are all monotone. Extensions that eliminate these latter two assumptions are left to the reader (see Exercises 7 and 8), but such extensions then yield proofs of the results stated in Chapter 1 as Theorems 1.3.1, 1.3.2, and 1.3.3.

Theorem 3.4.3. *Suppose that V is a monotone voting rule in the context of linear ballots and three or more alternatives, and that V satisfies the following:*

(1) Pareto: if everyone ranks a *over* b, *then* b *is a non-winner.*
(2) CIIA: if a *is a winner and* b *is a non-winner, and ballots are changed, but everyone who had* a *over* b *keeps* a *over* b *and vice-versa, then, in the new election,* b *is still a non-winner.*

Then V is resolute and there is a dictator.

Proof: For $X \subseteq N$ and $a, b \in A$, say that aXb if, for every profile **P** in which everyone in X has a over b on their ballots, $b \notin V(\mathbf{P})$.

Claim 1 (the existence lemma): Suppose there exists a profile **P** in which everyone in X has a over b, everyone not in X has b over a, and for which $a \in V(\mathbf{P})$ and $b \notin V(\mathbf{P})$. Then aXb.

Proof. Suppose, for contradiction, that aXb fails. Then we have a profile $\mathbf{P}'$ in which everyone in X has a over b, and yet $b \in V(\mathbf{P}')$. By monotonicty of V we can assume, in $\mathbf{P}'$, that everyone not in X has b over a. But now we can change the ballots in **P** so that they become identical to those in $\mathbf{P}'$ and, by CIIA, conclude that b is still a non-winner. This contradiction completes the proof of Claim 1.

Claim 2 (the splitting lemma): If aXb, and c is distinct from a and b, then for every partition of X into disjoint sets Y and Z (one of which may be empty), either aYc or cZb.

Proof. Consider the following profile P, in which every voter places all alternatives other than a, b, and c below these three in any order

[16] For a series of different proofs of Arrow's theorem, the reader can start with Fishburn (1970), Blau (1972), Barberá (1980), and Geanakoplos (1996).

whatsoever:

Y	Z	N–X
a	c	b
b	a	c
c	b	a

By Pareto, the winners are among a, b, and c. Because $a\text{X}b$, b is a non-winner. If c is a winner, then $c\text{Z}b$ by the existence lemma. If c is a non-winner, then a must be a winner (being the only alternative not yet ruled out). Thus, $a\text{Y}c$ by the existence lemma, and this proves Claim 2.

Notice that Pareto implies that $a\text{N}b$ for every a and b, but that it, together with CIIA, implies $a\emptyset b$ fails for every a and b (because, in particular, if everyone has b on top of his or her ballot, then b is the unique winner). It now follows from Proposition 3.4.2 that there is a voter i such that $a\{i\}b$ for every a and b. But this clearly means that the election winner is always the unique alternative that voter i has on top of his or her ballot. This completes the proof of Theorem 3.4.3.

□

The version of monotonicity that we need for the next result, dealing with social welfare functions, asserts that if a is not over b on the final list, and some voter who had a over b on his or her ballot interchanges a and b, then a is still not over b on the final list.

Theorem 3.4.4. *Suppose that V is a monotone social welfare function in the context of linear ballots and three or more alternatives, and that V satisfies the following:*

(1) *Pareto: if everyone ranks* a *over* b, *then* a *is over* b *on the final list.*
(2) *IIA: if* a *is over* b *on the final list, and ballots are changed, but everyone who had* a *over* b *keeps* a *over* b *and vice versa, then, in the new election,* a *is still over* b *on the final list.*

Then V is resolute and there is a dictator.

Proof: For $\text{X} \subseteq \text{N}$ and $a, b \in \text{A}$, say that $a\text{X}b$ if a is over b on the final list whenever everyone in X has a over b on their ballots.

Claim 1 (the existence lemma): Suppose there exists a profile **P** in which everyone in X has a over b, everyone not in X has b over a, and for which a is over b on the final list V(**P**). Then $a\text{X}b$.

Proof. Suppose, for contradiction, that $a\mathrm{X}b$ fails. Then we have a profile $\mathbf{P}'$ in which everyone in X has a over b, and yet a is not over b on the final list $\mathrm{V}(\mathbf{P}')$. By monotonicty of V, we can assume that everyone not in X has b over a in $\mathbf{P}$. But now we can change the ballots in $\mathbf{P}$ so that they become identical to those in $\mathbf{P}'$ and, by CIIA, conclude that a is still over b on the final list. This contradiction completes the proof of Claim 1.

Claim 2 (the splitting lemma): If $a\mathrm{X}b$, and c is distinct from a and b, then for every partition of X into $\mathrm{Y} \cup \mathrm{Z}$, either $a\mathrm{Y}c$ or $c\mathrm{Z}b$.

Proof. Consider the following profile $\mathbf{P}$, where every voter places all alternatives other than a, b, and c below these three in any order whatsoever:

Y	Z	N–X
a	c	b
b	a	c
c	b	a

By Pareto, a, b, and c are ranked above all other alternatives in the final list. Because $a\mathrm{X}b$, a is over b on the final list. If c is over b on the final list, then $c\mathrm{Z}b$ by the existence lemma. If c is not over b on the final list, then a is over c on the final list (because a is over b, and b is over c or tied with c on the final list). Thus, $a\mathrm{Y}c$ by the existence lemma, and this proves Claim 2.

Notice that Pareto not only implies that $a\mathrm{N}b$ for every a and b, but that $a\emptyset b$ fails for every a and b. It now follows from Proposition 3.4.2 that there is a voter i such that $a\{i\}b$ holds for every a and b. But this clearly means that the final list is identical to voter i's ballot. This completes the proof of Theorem 3.4.3. □

Because we know what monotonicity means in the context of a social welfare function, there is no need to define what it means in the context of a social choice function that satisfies transitive rationality.

Theorem 3.4.5. *Suppose that V is a monotone social choice function satisfying transitive ratonality in the context of linear ballots and three or more alternatives, and that V satisfies the following:*

(1) Pareto: if everyone ranks a *over* b *then* a *is over* b *on the final list.*
(2) IIA: if a *is over* b *on the final list, and ballots are changed, but everyone who had* a *over* b *keeps* a *over* b *and vice versa, then, in the new election,* a *is still over* b *on the final list.*

Then V is resolute and there is a dictator.

Proof: Because we are assuming transitive rationality, the social choice function V is really a social welfare function. Hence, Theorem 3.4.5 follows immediately from Theorem 3.4.4. □

3.5 Exercises

(1) [S] Suppose that N is an eight-element set and that G is a collection of subsets of N such that (*i*) $N \in G$, and (ii) for every $X \subseteq N$, if $X \in G$ and X is partitioned into sets Y and Z, then either $Y \in G$ or $Z \in G$. Give three different proofs that $\{i\} \in G$ for some $i \in N$.

(2) [S] Prove that a resolute voting rule that satisfies down-montonicity also satisfies monotonicity.

(3) [S] Prove that in the context of non-linear ballots and three or more alternatives, there are weak dictatorships that are manipulable.

(4) [S] Prove that in the context of non-linear ballots and three or more alternatives, there are weak dictatorships that are non-manipulable.

The following two exercises allow one to derive a version of Arrow's theorem for non-linear ballots from a version for linear ballots. Notice, however, that we are using MIIA (the version of IIA that incorporates monotonicity).

(5) [T] Suppose that V satisfies MIIA and that for every linear ballot **P**, $V(\mathbf{P}) = \{\mathrm{top}_i(\mathbf{P})\}$. Prove that for every ballot **P**′, $\mathrm{top}_i(\mathbf{P}') \in V(\mathbf{P}')$.

(6) Derive Arrow's theorem for non-linear ballots from the version for linear ballots.

(7) [T] Show that the assumption of monotonicity can be eliminated from Theorem 3.4.5 by establishing that if $b \notin V(\mathbf{P})$ for every profile **P** in which everyone in X has a over b on their ballots, and everyone else has b over a on their ballots, then aXb.

(8) [T] Do a version of Exercise 7 that eliminates the monotonicity assumption from Theorem 3.4.4 and Theorem 3.4.5.

(9) [S not T] Prove that if V satisfies Pareto and can be manipulated by a coalition, then it can be manipulated by a single voter.