2

The Debates Surrounding Social Choice

The question of how multiple, competing, goals can be reconciled is the foundation of a branch of political science termed "social choice theory" or "collective choice theory," the study of which can be traced to the writings of a collection of French mathematicians and philosophers of the late eighteenth century, most notably Jean-Charles de Borda and the Marquis de Condorcet. Condorcet in particular was focused on the search for truth in public discourse, and his most well-known thoughts on politics are found in his *Essai sur l'application de l'analyse a la probabilite des decisions rendues a la pluralite des voix* ("Essay on the Application of Mathematics to the Theory of Decision-Making"). In this work he proved his famous jury theorem, which says that if a group is choosing between two alternatives (e.g., *acquit* or *convict*) and if each individual member of the group is more likely than not to reach a correct decision, then the probability that a majority of the members of the group reach the correct decision is higher than the probability that any individual reaches the correct decision and increases as the size of the group increases. This is a positive result, as it shows that a simple majority vote does well at producing a correct outcome when there are two alternatives and many voters.[1]

However, Condorcet also realized that group choices frequently involve more than two alternatives, and he sought a similar result for these cases. In playing with possible distributions of voter preferences, Condorcet, in his essay, identified what is now a well-known conundrum.[2] *Condorcet's paradox*, or more simply, *the paradox of voting* famously illustrates a problem stemming

[1] Condorcet was also deeply committed to liberal causes, including universal suffrage and universal public education for both men and women (along with subsidies to agricultural families whose children attended public schools and could no longer provide farm labor). See Block (1998), pp. 1015–1018.

[2] Block (1998), pp. 1004–1006.

from majority rule in which pairwise voting over three or more alternatives can lead to intransitive (or cyclic) outcomes. The paradox goes as follows: suppose there are three individuals (Persons 1, 2, and 3), and suppose there are three alternatives to be voted upon (*A*, *B*, and *C*). Finally, suppose that Persons 1, 2, and 3 have the following preferences over the alternatives:

Person	Preferences
1	$A \succ_1 B \succ_1 C$
2	$B \succ_2 C \succ_2 A$
3	$C \succ_3 A \succ_3 B$

As we discuss in more detail later in the chapter, the notation $A \succ_1 B$ means that Person 1 prefers *A* to *B*. With this in hand, now suppose that the above individuals are asked to cast pairwise votes over the three alternatives. Then, by the votes of Persons 1 and 3, a majority vote between *A* and *B* would yield *A* as the winner, the votes of Persons 1 and 2 would give *B* a majority over *C*, and the votes of Persons 2 and 3 would yield *C* as the winner between *A* and *C*. Thus, majority voting in this case produces a *cycle* by deeming alternative *A* superior to alternative *B*, alternative *B* superior to alternative *C*, and alternative *C* superior to alternative *A*. Put another way, no alternative can be deemed "best" in terms of satisfying a majority of voters.[3] Situations like this in which the majority preference relation is cyclic on a set of three alternatives is often referred to as a *Condorcet cycle*. Conversely, an alternative that is majority preferred to every other alternative is referred to as a *Condorcet winner*.

Condorcet's work on voting procedures had little influence on his contemporaries but was later rediscovered by Duncan Black in a series of essays written in the 1940s and culminating in his book *Theory of Committees and Elections*.[4] This work is considered to be the beginning of modern social choice theory.[5] In the 1950s economist Kenneth Arrow independently rediscovered the ideas of Condorcet while working on his text *Social Choice and Individual Values*. This book, published in 1951 and for which Arrow received the Nobel Memorial Prize in Economic Sciences in 1972, presents Arrow's seminal "impossibility theorem." Put succinctly, Arrows impossibility theorem demonstrates that when voters have three or more options to choose from, then any voting system that meets certain minimal conditions of fairness and sensibility *will* fail to produce "rational" outcomes in some situations. More specifically, if the system satisfies the fairness conditions, it must necessarily succumb to instances in which it deems an alternative *A* weakly superior to an alternative *B*, that

[3] More strongly, the three alternatives are essentially indistinguishable from each other: they are identical in every way except for identities of the voters that support them. This example arguably represents the foundation of social choice because of its special combination of simplicity and intractability.

[4] Riker (1982), p. 2. Note that some people, including Charles Dodgson, rediscovered the work in the nineteenth century.

[5] Block (1998), p. 984; Black (1958) and Riker (1982), footnote 13, pp. 1–2.

alternative *B* weakly superior to another alternative *C*, and that alternative *C* strictly superior to *A*. In other words, many minimally democratic systems will in some situation produce an intransitive ordering of the alternatives similar to the cyclic outcomes that Condorcet identified nearly 200 years earlier.

Social choice–theoretic concepts such as the paradox of voting and results such as Arrow's theorem have had a tremendous impact on the study of politics. Perhaps the most influential and controversial interpretation of these results can be found in the late William Riker's famous book, *Liberalism Against Populism*, in which he used the paradox of voting and Arrow's Theorem to argue that any notion of a "popular will" is meaningless, that political decisions are in a continual state of disequilibrium, and that perhaps the only benefit of democracy is that it enables voters to throw bad politicians out of office. This argument in favor of what some have termed "democratic irrationalism,"[6] or the idea that the preferences of individuals cannot be amalgamated in any meaningful way because of the possibility of majority voting cycles, has had a broad impact, influencing both academics and policymakers. For example, in the 1989 Supreme Court foreword to the Harvard Law Review, Erwin Chemerinsky uses Arrow's Theorem as the cornerstone of his argument that democratic legislatures cannot "reflect the views of a majority in society."[7]

In defending the idea that democratic choices do (or at least can) reflect a nonvacuous popular will against attacks like Riker's, a strain of literature has arisen that attempts to evade the consequences of Arrow's Theorem. This literature argues that Riker and his followers are misguided because the social choice results themselves are of limited relevance. These results, it is argued, are either empirically irrelevant, based on incorrect underlying assumptions, or both. Gerry Mackie's nearly 500-page tome *Democracy Defended* focuses exclusively on discrediting Riker's arguments by attempting to demonstrate that virtually every published empirical claim of a paradox arising from Arrow's Theorem has been made in error, and moreover, that the conditions of Arrow's Theorem have "no descriptive or normative force of their own."[8]

Despite the lively intellectual debate described above, over the past quarter century social choice theory as a descriptive tool has fallen out of fashion with academics. Even its allies have characterized the social choice impossibility results as intellectual dead ends, and much of social science theorizing now uses a game-theoretic approach, wherein equilibrium existence is much easier to obtain.[9] More challengingly, the field has been criticized as being nonfalsifiable and useless as an empirical tool.[10] Arrow's result has been described as "an

[6] Gerry Mackie's term.
[7] Chemerinsky (1989), p. 79, quoted in Pildes and Anderson (1990), p. 2124.
[8] Mackie (2003), p. 156.
[9] However, one should note the subtle and cogent points raised by David Austen-Smith and Jeffrey Banks about the relationship between these types of approaches (Austen-Smith and Banks (1999), pp. 187–194).
[10] Green and Shapiro (1994).

abstract limit case" that "does not describe the real world,"[11] and some have gone so far as to claim that

> [o]riginating in profound misconceptions about the structure of public values, the nature of democratic politics, and the concept of rationality itself, social choice theory only muddies efforts to think clearly about democracy.[12]

Taken as a whole, while few authors dispute that results such as Arrow's Theorem provide *some* insight into what voting systems are (and are not) capable of accomplishing, the real-world relevance of these results is less well understood and accordingly more contested. The remainder of this chapter presents a brief and semitechnical introduction to two of the most well-known impossibility theorems: Arrow's theorem and the Gibbard-Satterthwaite Theorem. We then describe Riker's view of the impossibility theorems, as his interpretation is regarded by many to motivate the role of these theorems in mainstream political science. Finally, the chapter concludes with a discussion of the various criticisms that have been leveled at Riker's arguments and at the impossibility results themselves. We set up these debates as a prelude to the chapters that follow, in which we hope to convince the reader that *both* Riker and his critics are wrong. Specifically, we will demonstrate that the social choice results are of real-world relevance and that the aggregation problem – the dilemma of comparing and reconciling competing interests and goals – is simultaneously the defining problem of political science and the logical foundation for democratic governance.

2.1 THE ARROW AND GIBBARD-SATTERTHWAITE THEOREMS

Arrow's Theorem and the Gibbard-Satterthwaite Theorem are two impossibility results that are commonly interpreted as applying to systems of voting (or, slightly more generally, methods of preference aggregation). We will argue later that such an interpretation is narrower than necessary, but this interpretation suffices for the purposes of describing the results. Adopting this interpretation for the time being also facilitates our discussion of existing debates about the results and their relevance to democratic politics.

The Arrow and Gibbard-Satterthwaite Theorems each take a minimal set of normatively appealing criteria and then formally demonstrate that these criteria are internally inconsistent. That is, each of the results implies that it is impossible for *any* voting system to simultaneously satisfy the given set of criteria. Despite this commonality, the theorems otherwise appear at first to be quite different. For example, Arrow's Theorem concerns the ability of a voting

[11] Mackie (2003), p. 156.
[12] Pildes and Anderson (1990), p. 2213.

procedure to produce outcomes that are collectively "rational," where collective rationality is described in terms of a procedure satisfying various common sense-type properties. The Gibbard-Satterthwaite Theorem, on the other hand, concerns the ability of a procedure to be immune from strategic manipulation by voters or to not reward insincere voting behavior. Despite their differences, the two results are mathematically similar, and it has been demonstrated that the two results can be derived from a more general "metatheorem" on social aggregators, which we discuss in more detail in the following chapter.[13]

To present these results, some simple notation is needed. First, we will assume that there are two or more individuals seeking to make a collective choice. Second, we will assume that there is a finite set of alternatives, or policies, under consideration by the group. We will denote this set by X, and will assume (to begin with) that X contains at least three different alternatives. Third, we will assume that each person has his or her own preference ordering of the alternatives under consideration. This *preference relation* is denoted $\succ_i$ for person i. Thus, if person i likes alternative x more than alternative y, it is denoted $x \succ_i y$. People are presumed to be *rational* in the sense that their preference relation is transitive: if $x \succ_i y$ and $y \succ_i z$, then $x \succ_i z$. Finally, we will use the term ρ to describe the entire collection of individual preferences, a set of preference relations for each person in the group under consideration. Thus, if there are n people in our group, then $\rho = (\succ_1, \succ_2, \ldots, \succ_n)$.[14]

To make this concrete, suppose that we have two people in our group and that we have three alternatives under consideration, so that $X = \{x, y, z\}$. Suppose that Person 1 strictly prefers x to y and z and strictly prefers y to z (i.e., $x \succ_1 y \succ_1 z$) and Person 2 strictly prefers y to z and x and z to x (i.e., $y \succ_2 z \succ_2 x$). Then the preference profile ρ characterizes our group and the preferences of its members as follows:

$$\rho = \begin{pmatrix} x \succ_1 y \succ_1 z \\ y \succ_2 z \succ_2 x \end{pmatrix} \tag{2.1}$$

Now that we have described the collection of alternatives, people, and their preferences over the alternatives, we can begin to consider various ways of conceptualizing collective choice. We will consider two different types of mechanisms for generating a group choice. The first is termed a *preference aggregation rule*, and the second is a *choice function*. Arrow's Theorem concerns preference aggregation rules; the Gibbard-Satterthwaite Theorem concerns choice functions. We now briefly discuss each of these two representations of collective decision making in turn.

[13] Reny (2001) and Eliaz (2004).

[14] Throughout this part of the book we presume that individual preferences are strict, so that if $x \succ_i y$, then it must be the case that *not* $y \succ_i x$. This is solely for ease of exposition. Collective preference, as generated through a to-be-defined preference aggregation rule, need not be strict, so that it may be the case that both $x \succ y$ and $y \succ x$.

PREFERENCE AGGREGATION RULES. A preference aggregation rule takes a preference profile ρ as an input and generates a collective preference relation, $\succ$, over all alternatives. An arbitrary preference aggregation rule is denoted by f, so that $f(\rho) = \succ$ describes the "group's preferences" over the alternatives when the individual preferences are as described by ρ.

While we use the term "collective preference" to refer to $\succ$, it is important to note that $\succ$ need not be transitive. As mentioned earlier in our discussion of Condorcet's seminal contributions, the paradox of voting is based on a preference aggregation method (pairwise majority voting) that can generate cyclic collective preferences. A preference aggregation rule that always returns a transitive collective preference relation is commonly referred to as a *social welfare function*. An example of a preference aggregation rule f that is also a social welfare function is the method of Borda count.

The Borda count method, which we will denote by f_B, or $\succ_B$, works as follows. For each individual preference relation $\succ_i$, *each alternative x receives the number of points equal to the number of alternatives ranked below x* in $\succ_i$. The collective preference is then given by the ordering of the alternatives in terms of these points: an alternative x is weakly collectively preferred to another alternative y given a profile ρ if and only if x receives at least as many points as y does at ρ.

For example, consider the preference profile ρ described in Equation 2.1. As there are three alternatives under consideration, Borda count works as follows: an alternative that a voter ranks first receives two points, an alternative he ranks second receives one point, and an alternative he ranks third receives zero points. The social ranking is then the sum of these scores across individuals. Thus, given the ρ in Equation 2.1, x receives two total points (two from Voter 1 and zero from Voter 2), y receives three total points (one from Voter 1 and two from Voter 2), and z receives one total point (from Voter 2). It follows that Borda count ranks the alternatives $y \succ_B x \succ_B z$.

Preference aggregation rules are appealing because what they produce – a collective preference relation – allows a collective comparison of any pair of alternatives. This is useful if one believes that preference aggregation involves contingencies; in some cases the group may be forced to rank alternatives before knowing which of the alternatives will actually be feasible. Put another way, the notion of collective preference facilitates an analogy between individual and collective choice and, accordingly, collective and individual rationality. From a practical standpoint, however, true preference aggregation is often unnecessary or inefficient. That is, while there are strong theoretical arguments in favor of representing group decision making as a preference aggregation rule, it is rare that preference aggregation rules are needed or applied. Rather, many collective decision-making procedures return simply a final choice. Choice rules bear a greater verisimilitude to such institutions, and we now turn to these.

CHOICE FUNCTIONS. While a preference aggregation rule f takes a preference profile ρ and produces a collective preference ordering over the entire collection of alternatives under consideration, a *choice function* takes a preference profile and returns a single alternative as the final collective choice. Thus, while preference aggregation rules are admittedly a scholarly abstraction, a choice function is similar to an electoral rule: individuals submit their preferences and the choice function identifies the winner. We denote an arbitrary choice function by F, so that the final choice when the preference profile is given by ρ is denoted by $F(\rho) \in X$. We can also think of Borda count as a choice function, denoted by F_B. In this case, our function would return the alternative with the highest Borda score, breaking a tie by some arbitrary rule if need be.[15] Referring again to the profile ρ described in Equation 2.1, the Borda count as a choice function would simply select alternative y:

$$F_B(\rho) = y.$$

Of course, choice functions and preference aggregation rules are distinguished only by what they produce. Because choice rules return a single alternative and the individuals are presumed to have preferences over the set of alternatives, choice functions have been used extensively to consider the effect of collective choice mechanisms on individual incentives when voting and otherwise signaling their preferences. We return to this question later when we discuss the Gibbard-Satterthwaite Theorem, but it is important to consider for a moment the linkage between choice functions and preference aggregation rules. A choice function produces a single winner, while an aggregation rule produces a comparison between each pair of alternatives, represented by $\succ$. If it is the case that the aggregation rule is such that there exists some $y \in X$ such that $y \succ x$ for all $x \in X \setminus \{y\}$ and for all such x it is also not the case that $x \succ y$, then y is the uniquely top-ranked element of $f(\rho)$ and a natural "collective choice" given that aggregation rule. However, if $f(\rho)$ is such that two or more alternatives are tied as best, or if $f(\rho)$ is cyclic and there *is no best*, then the aggregation rule cannot produce a collective choice without more structure being placed on the collective decision-making process.

The theory we present in Chapters 5 and 6 provides an answer to the question of how to appropriately (or "legitimately") translate a cyclic collective preference relation into a single collective choice. As we will discuss throughout the next few chapters, this situation characterizes the heart of the debate about the political relevance of Arrow's Theorem in particular and the theory of social choice in general. Cyclic collective preference is potentially insufficient to unambiguously discern a collective choice and to construct a coherent representation of a "collective will." By bridging this divide between the output of "pure aggregation" (of preferences or other criteria) and the selection of a

[15] How one breaks ties in choice functions can have very important consequences for both individual and collective behavior. However, these issues, and tie breaking in general, are not relevant to our purposes in this book.

final choice, our theory provides a reconciliation of the inputs of the aggregation problem and the instrumental requirement that one policy ultimately be chosen.

Given the centrality of the problem of cyclic collective preferences, a natural next question is why one would adopt a preference aggregation rule that might ever produce such a thing. That is, are there reasons that one might choose an aggregation method that might not yield an unambiguous "best" outcome? Arrow's Theorem provides an affirmative answer to this question. In a nutshell, any minimally democratic aggregation procedure must encounter some situations in which it fails to produce a coherent (or, perhaps, "well-ordered") collective preference. In the following section, we turn to a more precise definition and discussion of how the theorems of Arrow and Gibbard-Satterthwaite characterize minimally democratic preference aggregation rules and choice functions, respectively. Before this, however, we briefly discuss the question of what preference aggregation rules and choice functions must take as "inputs." In social choice terms, this issue is described as the *domain* of the aggregation rule or choice function.

Preference Domains

Recall that preference aggregation rules and choice functions are each presumed to take preference profiles as their inputs. The only distinction between the two concepts is what they produce upon receiving a preference profile.[16] The set of preference profiles for which a preference aggregation rule or choice function (or more simply, a "rule") is defined is referred to as the rule's *domain*.

A rule that is capable of considering (i.e., defined for) all possible preference profiles is said to satisfy *unrestricted domain*. In other words, as long as a choice function always returns a choice (or, respectively, a preference aggregation rule always returns a collective preference relation), it satisfies unrestricted domain. Of course, requiring that a rule satisfy unrestricted domain does not imply that every preference profile is possible. Accordingly, unrestricted domain is in reality simply a technical condition that is satisfied by any well-defined rule.

Some scholars have argued in both substantive and normative terms that unrestricted domain concerns whether the rule can (or should) restrict the preferences that individuals may have.[17] These arguments, it should be noted, are completely beside the point. Preference aggregation rules and choice functions are necessarily abstract constructions and to assert that a real-world instantiation of such a rule does not satisfy unrestricted domain is either inaccurate

[16] It is important to note again at this point that we are using the term "preference" here simply for the purposes of illustration. One of our main points is that the social choice results we discuss apply to any aggregation problem, regardless of the substantive nature or context of the inputs.

[17] Or, perhaps, whether a rule can or should restrict the preferences that individual may claim to have. We return to this point, and the previous arguments we allude to here, in Section 3.2.

or an assertion that there is some situation in which the rule "does nothing." Deferring discussion of the possible inaccuracy of this statement for the moment, it is important to consider how nonsensical "doing nothing" is in this context. It is impossible for the output of a real-world rule to "not be defined." There is always something that happens after a real-world rule is given a set of inputs (e.g., a preference profile). Generally, the phrase "do nothing" is used to describe what happens when the rule makes no change to the prevailing policy. Regardless, some policy will indeed prevail after the rule receives its inputs. Whatever this policy is *defines the rule*. In short, unrestricted domain is truly a regularity condition for the purposes of analysis.

An inaccuracy that offers an explanation for the debates about unrestricted domain concerns a traditional divide between theory and empirics. As mentioned earlier, in no way should the assumption of unrestricted domain be taken to imply anything about the presumed frequency or possibility of the rule in question *observing* every preference profile as an input. However, this is an empirical question first and foremost. As we discuss in more detail in Section 3.2, there are misunderstandings about exactly how much purchase one gets from restricting the set of preference profiles that can be observed. Even more important, these arguments have almost always ignored the fact that aggregation in political institutions frequently involves inputs other than preferences. Ultimately, though, the question at hand in such empirical debates is not about whether a rule satisfies unrestricted domain. Rather, these debates should be interpreted in terms of the degree to which one should be concerned about satisfaction of the other democratic criteria, to which we now turn.

Arrow's Theorem

Arrow lays out four simple axioms that he argues any reasonable aggregation rule should satisfy. He then proves that these axioms are incompatible with each other, that no rule can simultaneously satisfy all four.[18] In so doing, his result implies that *any* aggregation rule – regardless of what is being aggregated or for what purpose – must violate at least one of these axioms. Put another and more specific way, every democratic institution, be it electoral, legislative, administrative, or judicial in character, violates at least one of these axioms. We now define each of these four axioms – Pareto efficiency, independence of irrelevant alternatives, transitivity, and no dictator – in turn.

Pareto Efficiency

We begin with the observation that perhaps one of the least demanding requirements a group would seek to impose on its voting system is that a group decision

[18] It should be noted that Arrow's original theorem (Arrow, 1951) used the axioms of *monotonicity* and *non-imposition* instead of the Pareto condition described below. The more common version of the theorem presented here (Arrow, 1963) replaces those axioms with Pareto efficiency and is a stronger result, because it uses weaker conditions.

be minimally responsive to the preferences of the members of that group. Arrow captures the notion of minimal responsiveness with the condition of *Pareto efficiency*. An aggregation rule f is *Pareto efficient* if whenever *every* individual i strictly prefers x to y, then our aggregation rule f generates a collective ranking of the alternatives that ranks x strictly higher than y. This condition rules out aggregation rules that, for example, always rank $x \succ y$, regardless of the group members' x, y preferences.

Independence of Irrelevant Alternatives

The second condition Arrow required of an aggregation method is that it should not consider "irrelevant" alternatives when generating a ranking between two other alternatives. Specifically, the group members' preferences between alternatives c and d should not affect how the group decides between two different alternatives, a and b. This property is captured by Arrow's second condition, which is termed *independence of irrelevant alternatives*.

An aggregation rule f is *independent of irrelevant alternatives* (IIA) if for any two *different* profiles, ρ and ρ' in which each individual's x, y ranking under ρ agrees with their x, y ranking under ρ',[19] then the collective ranking of x and y ranking generated by $f(\rho)$ should agree with the collective ranking of x and y ranking generated by $f(\rho')$. In other words, if something alters people's preferences only about alternatives other than x and y, the collective ranking of x and y should remain the same.

As we discuss in more detail in Section 3.3, IIA is the most conceptually difficult of Arrow's conditions. This difficulty largely stems from the fact that it is a condition that applies *across* different preference profiles. For example, Pareto efficiency and transitivity are intraprofile conditions; it is possible to determine a violation of either of these axioms by considering a single preference profile. This is not the case with IIA-to determine that an aggregation rule violates IIA requires that one compare the output of the rule for at least two different preference profiles.[20]

While IIA is an abstract condition, an equivalent formulation is this: if any one individual's ranking of a particular alternative under consideration (call it z) changes, then this change alone should not affect how the *group* decides between two other alternatives, x and y. It is important to pause for a moment and consider this reformulation. If an aggregation method f violates IIA, then there is a pair of preference profiles, ρ and ρ' that (1) differ only with respect to the preferences of *one* individual, (2) do not differ at all with respect to *any* person's ranking of two alternatives x and y, and (3) at these two profiles f generates different collective rankings for x and y. Thus, it should be clear that violating IIA opens up the possibility for strategic manipulation

[19] That is, for each individual i, $x \succ_i y$ if and only if $x \succ_i' y$.

[20] This interprofile characteristic is shared by the no dictator axiom discussed later. However, no dictator is quite transparent in both its implications and, relatedly, which aggregation rules violate it.

of the aggregation process: a single individual may in some cases have the opportunity to alter the ranking of two alternatives simply by misrepresenting his or her preferences about some other alternative. This possibility is at the heart of the Gibbard-Satterthwaite Theorem, which we discuss later in this section. For now, though, consider the following example to make the concept behind IIA more concrete. In so doing, we will also illustrate how and why the Borda count procedure violates IIA.

Consider the following two profiles, ρ and ρ':

$$\rho = \begin{pmatrix} x \succ_1 y \succ_1 z \\ y \succ_2 z \succ_2 x \end{pmatrix} \qquad \rho' = \begin{pmatrix} x \succ_1 z \succ_1 y \\ y \succ_2 x \succ_2 z \end{pmatrix} \qquad (2.2)$$

As discussed earlier, Borda count (denoted again by f_B, or $\succ_B$), when applied to profile ρ, collectively ranks y above both x and z and ranks x above z:

$$y \succ_B x \succ_B z.$$

Meanwhile, at ρ', Borda count ranks x above y and z and ranks y above z:

$$x \succ'_B y \succ'_B z,$$

because at this profile x receives three combined points, y receives two, and z receives one. Note now that if we look solely at the two individuals' rankings of x and y, these two profiles look identical: voter 1 prefers x to y ($x \succ_1 y$) under both ρ and ρ', and Voter 2 prefers y to x ($y \succ_2 x$) under both ρ and ρ'. However, $f_B(\rho)$ generates $y \succ_B x$, and $f_B(\rho')$ generates $x \succ'_B y$. Accordingly, Borda count violates IIA.[21]

The fact that Borda count violates IIA is in a sense the basis of some of the more sustained critiques of the axiom as a desideratum of aggregation methods, and we discuss this debate in much greater detail in Section 3.3. Now, however, we turn to the third axiom, transitivity.

Transitivity

Arrow's third condition, *transitivity*, focuses on the ability of a preference aggregation rule to generate an unambiguous winner, or collection of winners, if there is a tie. As discussed earlier when comparing preference aggregation rules and choice functions, an aggregation rule that generates the social ranking $x \succ y$, $y \succ z$, and $z \succ x$ is not particularly useful to a group seeking to collectively choose one alternative from among x, y, and z. An aggregation method that does return such a cyclic relationship is said to *cycle*. In particular, an aggregation method that cycles may not provide an unambiguously "best" alternative when it returns a cyclic relationship between some or all of the alternatives.

An aggregation rule f is *transitive* if it always produces a transitive ordering of the alternatives. Thus, if f produces an ordering in which $x \succ y$ and

[21] Indeed, IIA is the only one of Arrow's four axioms that the Borda method violates.

$y \succ z$, then it must also be the case that $x \succ z$. This condition guarantees that the social ordering generated by f satisfies the same rationality condition as the individual preference orderings it was constructed from and that it cannot cycle. Moreover, it ensures the existence of an alternative (or collection of alternatives) that are not ranked strictly lower than anything else. As we discuss in more detail in Section 3.6, one can defend the desirability of this axiom from a number of perspectives, most notably the degree to which aggregated "social" preferences can be thought of as equivalent to individual preferences. In this light, Arrow's Theorem indicates important normative and logical concerns with anthropomorphizing groups when discussing group decision making. Deferring this discussion, however, we now discuss the fourth and final of Arrow's axioms, no dictator.

No Dictator

Arrow's final axiom, *no dictator*, concerns the responsiveness of the preference aggregation rule to the preferences of more than one person. An aggregation rule is *dictatorial* if there is one particular voter whose individual preferences always determine the social preference ordering, irrespective of the preferences of the other voters. Formally, this condition says that there exists one voter i, so that every time $x \succ_i y$, the aggregation rule f produces a strict ranking $x \succ y$. An aggregation rule f satisfies *no dictator* if it is not dictatorial. We discuss defenses of the no dictator condition in more detail in Section 3.5, but it is useful at this point only to note how weak this axiom is. In particular, a dictatorial aggregation rule is *completely independent* of all of the inputs to the aggregation problem except one.[22] If there is even one preference profile and one pair of alternatives at which the aggregation does not exactly match a given voter's preference ordering, then that voter is not a dictator under the rule.[23]

With unrestricted domain and the four axioms of Pareto efficiency, IIA, transitivity, and no dictator defined and described, we are now ready to state Arrow's Theorem.

Theorem 1 *(Arrow, 1950, 1963). With three or more alternatives, any aggregation rule satisfying unrestricted domain, Pareto efficiency, IIA and transitivity is dictatorial.*

[22] This claim relies on individuals having strict preferences. If a dictator has weak preferences (in which some alternatives are tied in his estimation), then the output of the dictatorial aggregation rule is not defined by the dictator and may (or may not) be generated by considering other individuals' preferences.

[23] More specifically, an aggregation rule can satisfy no dictator and nonetheless always return one given voter's most preferred alternative as the collectively most preferred alternative. This point is important both when contrasted with the definition of a dictatorial choice function below and, more generally, when considering the previously discussed linkages between aggregation rules and choice functions.

Arrow's Theorem then tells us that if a group wishes to design a preference aggregation rule that is Pareto efficient, transitive and independent of irrelevant alternatives, and if we place no restrictions on the preferences that individuals may have, then the rule must grant all decision-making authority to a single individual. Thus, any aggregation rule that is not dictatorial *must* violate transitivity, Pareto efficiency, or IIA. And practically speaking, it will violate either transitivity or IIA because the only nondictatorial aggregation rules that are ruled out by the addition of Pareto efficiency are those rules that are either null (generate a tie over all alternatives) or inverse dictatorships.[24] This extension of Arrow's Theorem to non-Pareto efficient rules was proved by Wilson (1972) and is known as Wilson's Impossibility Theorem.

The Gibbard-Satterthwaite Theorem

Our second impossibility theorem, proved independently by Gibbard (1973) and Satterthwaite (1975), differs from Arrow's Theorem in several important ways. First, it concerns choice functions rather than preference aggregation rules (i.e., rules that produce a single winner, as opposed to a social ordering of the alternatives). Second, the Gibbard-Satterthwaite Theorem does not assume that the rule is given a "true" preference profile as an input. Rather, it considers rules (e.g., voting systems) that take reported preferences, (e.g., individuals' ballots) as an input. The focus, in this case, is then on whether there are choice functions that can be relied upon to elicit truthful inputs (e.g., "sincere" ballots). In slightly different terms, the Gibbard-Satterthwaite Theorem considers whether and how a choice function might be implemented so as to make collective decisions when the preference profile must be elicited from individuals with an interest in the collective decision itself.

Formally, Gibbard and Satterthwaite consider what is referred to as the *strategy-proofness* of a choice function. A strategy-proof choice function entirely negates any gains from insincere behavior by any single voter. Consider two preference profiles, ρ and $(\succ'_i, \rho_{-i})$, that are as follows:

$$\rho = (\succ_1, \dots, \succ_i, \dots, \succ_n)$$

is a "true," or "sincere," preference profile, and

$$(\succ'_i, \rho_{-i}) = (\succ_1, \dots, \succ'_i, \dots, \succ_n)$$

is a profile that differs from ρ only in that Voter i reports the "insincere" or "incorrect" preferences $\succ'_i$, as opposed to his true preferences $\succ_i$. A choice function F is *strategy-proof* if, in every situation, ρ, F never chooses an alternative at $(\succ'_i, \rho_{-i})$ that Voter i strictly prefers to the outcome it selects

[24] Wilson's Theorem additionally requires the very weak axiom of non-imposition.

at ρ. Formally, for a strategy-proof F, it is the case that for every ρ and every $i \in N$:

$$F(\rho) \neq F(\succ'_i, \rho_{-i}) \Rightarrow F(\rho) \succ_i F(\succ'_i, \rho_{-i}).$$

In other words, F being strategy-proof implies that no voter can ever strictly benefit by claiming to have preferences that are different than what they actually are (or more specifically, by claiming preferences $\succ'_i$ when his true preferences are $\succ_i$). Put less formally, honesty is always a "good policy" when voting or otherwise submitting information to a strategy-proof choice function.

Note that there is a class of very simple choice functions that are strategy proof. One could simply choose a single voter and choose whatever that voter reports as his or her most preferred alternative. That voter can never strictly gain from misreporting his or her preferences. Similarly, none of the other voters can affect the chosen alternative by what they report, so they too have no incentive to report something other than their true preferences. Such choice functions are referred to as *dictatorial*. Note that the definition of "dictator" used in this theorem is modified slightly from our previous definition to accommodate the fact that we are considering choice functions: here, a choice function F is dictatorial if it always generates a collective choice that is the dictator's top-ranked alternative. Gibbard and Satterthwaite demonstrate that, if at least three different voting outcomes are possible, these choice functions are the *only* ones that are strategy proof. In other words, there is no nondictatorial procedure that is strategy proof.

Theorem 2 *(Gibbard, 1973; Satterthwaite, 1975) With unrestricted domain and the possibility of three or more voting outcomes, any strategy-proof choice function is dictatorial.*

From the perspective of choice functions as representing voting systems, the Gibbard-Satterthwaite Theorem proves that the possibility of strategic voting, or voting against one's true preferences, is endemic to *every* nontrivial voting system. While the scope of this result is surprising (in that it tells us that there is *no* nondictatorial voting system that is nonmanipulable), it should not be surprising that individuals frequently have incentives to cast insincere ballots in elections. In the plurality system frequently used in elections in the Unites States, for example, supporters of third-party candidates often have a perceptible incentive to vote for their favorite major party candidate because a vote for a third party may be considered a "wasted vote."

The profundity of the Gibbard-Satterthwaite Theorem is more easily seen when one adopts a more general interpretation of choice functions than as mere voting systems. In particular, when one abstracts from individual preferences and conceives of the "preference profile" as a profile of objective information about different criteria that the choice function is designed to use when selecting a final outcome, strategy proofness can be more easily reinterpreted as requiring the choice function to respond to the criteria in a way that is, for lack of a better word, faithful to each of them. In other words, nothing in the Gibbard-Satterthwaite Theorem restricts its applicability to voting or electoral systems.

Implications of the Theorems

Having described both Arrow's Theorem and the Gibbard-Satterthwaite Theorem, it is natural to ask: What do these theorems tell us about democratic processes? After all, the results themselves are purely mathematical statements and, indeed, from a technical standpoint, each is a relatively simple and straightforward application of formal logic. Nonetheless, the stark clarity provided by this abstraction imbues the conclusions with the rarefied quality of truth. Simply put, the theorems are indisputably *correct*. Accordingly, it follows that one's perception of the value, or lack thereof, of these results depends entirely on how one chooses to interpret them.

For the remainder of this chapter we discuss several strands of literature that argue that despite the optimistic name Arrow gave to his theorem – the "General Possibility Theorem" – the social choice enterprise as pioneered by Arrow is a fundamentally hopeless endeavor.[25] Although these literatures are at odds with each other, and indeed one arose as a refutation of the other, we argue that they both take a similar conception of successful democratic choice. One strand argues that democracy, conceived of as the amalgamation of the preferences of members of society, is impossible because of the social choice results; another argues that democracy is possible because these results are not likely to be of real-world relevance. We argue that both approaches are wrong – that democracy is meaningful precisely *because of* the far-reaching relevance of the impossibility theorems.

2.2 RIKER AND THE ARBITRARINESS OF DEMOCRATIC CHOICE

Early on in *Liberalism Against Populism* in a chapter titled "Different Choices from Identical Values," William Riker discusses various voting systems used throughout the world and observes that, for many profiles of preferences, different voting systems will yield different outcomes. Each of the systems is a sensible one (otherwise we might not expect it to be widely used), and the fact that these systems yield different outcomes is not entirely surprising (otherwise we might expect that a nation's choice of electoral system is not a particularly important or meaningful choice). At the same time, the fact that different fair and sensible procedures can produce different outcomes when applied to the same set of voters troubles Riker because it implies that real-world voting outcomes cannot be regarded as true and accurate amalgamations of voters' preferences.

[25] Sen (2012), p. 263.

Taking this fact as a starting point, Riker goes on to make the stronger claim that even if an unambiguously "best" voting system was agreed upon, the aforementioned impossibility theorems prove that we still cannot take its output as any meaningful reflection of the popular will. As discussed earlier in Section 2.1, Arrow's Theorem implies that if the (nondictatorial) rule satisfies the "fairness conditions" of Pareto efficiency and independence of irrelevant alternatives, then it must produce intransitive outcomes; essentially, it must cycle.[26] Riker argues that this conclusion implies that any outcome produced by a "fair" electoral system is meaningless.

Second, the Gibbard-Satterthwaite Theorem implies that this agreed-upon system faces a potentially even bigger hurdle than the possibility of cyclic outcomes: it would not be able to elicit the truthful preferences of voters and would always be susceptible to situations in which voters faced a strategic incentive to misrepresent their preferences by casting "insincere" ballots. Thus, real-world voting outcomes cannot be regarded as accurate amalgamations of voters' preferences because the voting systems themselves have no way of eliciting what those preferences actually are. Riker argues that because social amalgamations of individuals' preferences are meaningless, a populist conception of democracy in which voters' preferences are translated into social outcomes, such as through a direct vote, is "absurd."[27] Rather, the best we can hope for is what Riker terms a liberal democracy, in which voters may not see their wishes translated into outcomes by their leaders but are free to vote their leaders out of office.

This is clearly a provocative argument, and it is an argument that has pervaded the study of social choice theory to such an extent that many scholars equate Riker's counterdemocratic interpretation of the impossibility theorems with the field of social choice itself.[28] In this sense, social choice theory as a field has come under fire as undermining the normative appeal of democracy itself. It is not surprising, then, that Riker's work has spurred a widespread effort by democratic theorists to defend democratic ideals in the face of this attack. If one takes Riker's argument seriously and takes his interpretation of the social choice results to be correct, then an obvious line of defense is to discredit the impossibility theorems themselves. This is an avenue that many of these critics have taken, and we discuss these critics in the sections that follow.

Another line of criticism has come from the positive political theory community, the establishment and development of which Riker himself played an

[26] It should be noted (and Riker notes) that transitivity is a sufficient condition to ensure that an aggregation rule does not cycle, but it is not a necessary condition. A necessary and sufficient condition is that the rule be *acyclic*. When Arrow's Theorem is extended to include this broader class of rules, his dictator condition is weakened to the existence of an individual who can veto certain decisions.

[27] Riker (1982), pp. 238–239.

[28] See Shepsle and Bonchek (1997), Mackie (2006)'s "Reception of Social Choice Theory by Democratic Theory," and the citations therein.

instrumental role in while at the University of Rochester in the latter half of the 20th century.[29] These critics argue that Riker's argument contains several logical inconsistencies. Coleman and Ferejohn (1986) summarize these criticisms succinctly, noting that two crucial parts of Riker's argument are, first, that all (populist) democratic procedures produce *arbitrary* outcomes, and second, that constitutional limitations on the power of elected officials produce *less* arbitrary outcomes. The problem with the former claim is that it is not true; many authors have shown that certain institutional constraints can shrink the set of feasible policy outcomes considerably.[30] The problem with the second claim is that Riker provides no theoretical ground for the argument that liberal institutions produce outcomes that are less arbitrary than do populist institutions.[31] While we agree wholeheartedly with these critiques of Riker's argument, we omit a further discussion because our own criticism of Riker's argument takes a different tack. That is, while we agree that Riker's interpretation of the impossibility results misses his intended target for various reasons, we believe that a different interpretation of the social choice results provides a more effective refutation of Riker's conclusions.[32]

2.3 MACKIE'S DEFENSE OF DEMOCRACY

In *Democracy Defended*, Gerry Mackie argues that by rejecting populist notions of democracy, Riker rejects democracy itself because "[w]hat almost everyone means by democracy is what Riker calls populist democracy."[33] In defending democracy against Riker's particular school of thought, Mackie aims to accomplish a single task: to demonstrate "the possibility of the accurate and fair amalgamation of opinions and wants."[34] And because Riker's argument hinges in large part on the impossibility theorems of Arrow and Gibbard and Satterthwaite, this task is largely undertaken by an exhaustive denouncement of nearly every claim made by those authors. Thus, if these results are shown to have little or no normative force of their own, Mackie argues, then fair preference amalgamation is possible, and then Riker is wrong. How one moves from the *possibility* of fair amalgamation to the *implementation* of fair amalgamation is a different question but one that we will see is not particularly difficult to resolve given Mackie's argument.

29 Amadae (2003), p. 169.
30 See Shepsle (1979), Banks (1985), Miller (1977), Miller (1980), and McGann (2006), among others.
31 Coleman and Ferejohn (1986), p. 8.
32 In colloquial terms, one can describe many of the prior positive political theory critiques of Riker's arguments as "yeah, but..." objections, whereas we believe our critique to be of a more assertive "no" variety.
33 *Ibid.*
34 *Ibid.*

In an attempt to negate (or, perhaps, neuter) the impossibility theorems Mackie, point by point, argues that each criterion of fairness or sensibility used in the theorems is in some way misguided as a desideratum of preference amalgamation. Arrow's Theorem tells us that with an unrestricted preference domain, every aggregation rule that is Pareto efficient, independent of irrelevant alternatives, and transitive is a dictatorship. Mackie argues that there is no reason to think of unrestricted domain, Pareto efficiency, independence of irrelevant alternatives, or transitivity as necessarily desirable properties of voting rules. Therefore, there may be many sensible and good procedures that violate one or more of these conditions and that are nondictatorial.

More specifically, what is wrong with the axioms? While Mackie focuses in large part on independence of irrelevant alternatives and unrestricted domain, Mackie's critique of unrestricted domain is arguably the heart of his argument. More than half of the book's nearly 500 pages is dedicated to the notion of the Condorcet cycle (i.e., the "paradox of voting" discussed earlier). Specifically, Mackie attempts to debunk every empirical claim of a real-world preference cycle, many of which were originally forwarded by Riker. The possibility of cycles forms the basis of Riker's argument that democratic politics is meaningless, as a majority preference cycle occurs only if pairwise majority voting generates an intransitive outcome. If cycles do not occur in practice, then pairwise majority voting is capable of yielding an unambiguously best outcome consistent with the majority will. Accordingly, Mackie argues, the absence of majority rule cycles implies that a coherent and meaningfully populist conception of democracy is possible. Mackie's principal "defense" of democracy, then, involves demonstrating that majority preference cycles do not occur in practice.

Establishing that majority rule cycles do not occur in practice requires that one demonstrate that the realized distributions of individuals' preferences in society are (or at least tend to be) "restricted" and possess a common structure so as to yield a Condorcet winner. Put another way, one must show that, in a specific sense, there is never too much preference heterogeneity or diversity. Put another way, preferences must be similar enough to guarantee that there is a "best" outcome in the sense that it is preferred to every other alternative by a majority of voters. A standard restriction that implies the existence of such an alternative is that preferences satisfy a *single-peakedness condition*. We return to this condition later, but for now it can be described as requiring that the collection of alternatives essentially differ from each other along one identifiable dimension (e.g., policies can be ordered along a liberal-conservative axis, each individual has a favorite spot (or "ideal point") on this axis, and each individual dislikes policies that are increasingly liberal or conservative, moving away from the individual's ideal point). Mackie (and others, to be discussed later) argue that there are many reasons to think that peoples' preferences are restricted in one of these ways. Essentially, these reasons come down to the fact that people "live in the same world and have similar interests in that world;

for example most prefer prosperity to torture of kittens to suicidal nuclear war."[35] If preferences are sufficiently homogenous, then Condorcet cycles may not exist, and the practical import of Arrow's condition of unrestricted domain is negated. This is important because Arrow's Theorem implies only that an aggregation rule satisfying Pareto efficiency, IIA, and no dictator will violate transitivity at *some* configuration of preferences. If a rule produces undesirable outcomes only in situations that will never occur in practice (i.e., preference profiles that do not occur empirically), then, at least in practical terms, it is arguable that such a violation should not be viewed as a criticism of the rule.

As noted earlier, the proofs of the theorems of Arrow and Gibbard and Satterthwaite are closely related. For example, the proof of the Gibbard-Satterthwaite Theorem also leverages the presumption of unrestricted domain to show that any nondictatorial voting rule offers individuals an incentive to strategically misrepresent their preferences. As with Arrow's theorem, the conclusion is that this incentive exists "at some preference profile." If one posits that some preference profiles do not occur in practice, then the Gibbard-Satterthwaite Theorem 's conclusion may be irrelevant from an empirical standpoint. In particular, there are nondictatorial choice functions that induce no incentive for strategic misrepresentation if (stated and) revealed preferences can be restricted so as to not admit a cyclic majority preference relation.

Taken as a whole, refuting the relevance of unrestricted domain can lead to the conclusion that not only does a majority will exist but also that there also exists a voting system that can elicit it: majority voting over all pairs of alternatives. Mackie's interpretation of the Arrow and Gibbard-Satterthwaite Theorems can thus be summarized as follows. Attempts to directly and meaningfully translate individuals' preferences into a collective choice will not necessarily fail because, in general, majority preference cycles do not exist. In general, there exists an outcome that reflects the "majority will," and moreover, when such an outcome exists, many voting systems will be able to discover it.

Of course, the assertion that majority preference cycles will *never* occur empirically is ridiculous. Mackie acknowledges this but asserts that they are either rare or inconsequential. More important, perhaps, Mackie suggests that when a cycle does exist, a "cyclebusting" voting rule such as the Borda count or plurality rule can always be used.[36] While such a rule does not remedy the fact that any alternative chosen by the group will be viewed by a majority of voters as being inferior to some other alternative, Mackie argues that the rule will provide a measure of stability to the political process and will, hopefully, produce an outcome with a broad degree of support.[37]

[35] *Ibid.*

[36] *Ibid.*

[37] While our goal here is not to refute Mackie's arguments per se, it is important to note that Mackie's suggestion here is unsatisfactory for several reasons. Most important, as described earlier, the objections of Riker were partially founded on the fact that the choice of rule will in effect be a choice of the outcome in such situations (e.g., when there is a cyclic majority preference relation), there is no guarantee that any two "cyclebusting" rules such as Borda count or plurality rule will return the same choice. Accordingly, to argue that one should use a cyclebusting rule to deal with majority preference cycles, one is forced to suggest which rule should be used or at least the rule that should be used by the group to choose which rule should be used to select which policy should be chosen.

2.4 ADDITIONAL REBUTTALS BY DEMOCRATIC THEORY

Although Mackie provides what may be the most comprehensive critique of the countermajoritarian interpretation of social choice, other democratic theorists have tackled Riker's objections from different angles.

Deliberation and Structuring Preferences

One strain of democratic theory, exemplified in work by Habermas (1987), Miller (1992), and Dryzek and List (2003),[38] argues that the process of *deliberation* provides a route around the dilemmas raised by the impossibility theorems. Similar to Mackie, these authors take aim at Arrow's condition of unrestricted domain. However, unlike Mackie, who argues that the condition is *prima facie* incorrect in most situations, these authors argue that the deliberative process works to alter individuals' preferences in such a way so as to induce "preference structuration," or change. That is, as opposed to positing that preferences are initially restricted so as to yield an unambiguous "collective preference," this line of thought suggests that deliberation generates (or, "structures") individual preferences so as produce such a collective preference.

Dryzek and List (2003) provide one particularly clear account of the process of structuration, whereby deliberation produces consensus at least on a single, underlying, and shared dimension of conflict, if not on which alternative should be chosen. When this type of structuration occurs, the resulting preferences can satisfy the single-peakedness condition described earlier, and pairwise majority voting will not be susceptible to cycling. Thus, these authors argue, the impossibility results can be reconciled with "populist style" democratic voting when democratic procedures have a deliberative aspect.

Rejectionist Critiques

Another group of scholars, whom Mackie terms the "rejectionist democrats,"[39] argues that social choice theory as an intellectual endeavor is meaningless on its own merits and, more provocatively, that both the substance and ensuing interpretations of the impossibility theorems pose a threat to democracy itself.

[38] *Ibid.*

[39] *Ibid.*

Pildes and Anderson (1990) offer a critique in this vein, beginning their attack on social choice theory by asking: "...does the formal logic of social choice theory truly compel us to abandon the search for collective decision-making processes that are both fair and rational; must we relinquish the effort to find meaning in our collective actions?"[40] To prove their assertion that the answers to both questions are "no," they critique what they claim are the underpinnings of social choice theory: in particular, the notion that individual preferences, the inputs to the Arrovian aggregation rule, are "consistent" in the sense of being transitive. This is because to suppose that individuals have transitive orderings of the alternatives under consideration is, in Pildes and Anderson's words, to suppose that individuals seek "the maximization of some single value..."[41] And this cannot be the case; individuals seek to maximize multiple, potentially competing, values, and "as individuals actually experience these values, they do not rest on a single scale and cannot be reduced to comparisons along a single, shared dimension."[42] Without transitive inputs there is no reason to expect collective choices to be transitive, and thus, they argue, Arrow's Theorem is seriously misguided to the point of being totally meaningless.

While the logic of Pildes and Anderson's argument is fatally flawed,[43] we are nonetheless in complete agreement with many of their claims. Similar to them, we believe that the incommensurability of values may lead to irresolvable conflicts and that these conflicts pose particular challenges to democratic governance.[44] In the face of these challenges, we similarly see the role of democratic politics as being, in part, "to involve not just choices but the reasons behind those choices..."[45] However, unlike Pildes and Anderson, we argue that it is *precisely* the impossibility theorems that give import to these claims. After all, Arrow's Theorem tells us that, in the face of these irresolvable conflicts, democratic procedures must do more than simply select a best outcome, as such an outcome will often not exist. In such cases, it follows, democratic procedures must legitimate their choices on a basis other than being the unambiguous product of popular will.

To claim, as Pildes and Anderson do, that the impossibility theorems are irrelevant to our understanding of democratic politics because a richer theory of choice is needed is to ignore an important selection effect. In particular, arguing that the impossibility theorems are irrelevant because "...social norms and institutional rules and practices are critical to enable democratic outcomes to be

[40] Pildes and Anderson (1990), p. 2127.
[41] *Ibid*, p. 2214.
[42] *Ibid*.
[43] *Ibid*.
[44] *Ibid*, p. 2166.
[45] Pildes and Anderson (1990), p. 2166, and quoted in Mackie (2003), p. 35.

meaningful or coherent..."[46] misses the most crucial point: these results precisely and elegantly indicate *why* such norms, rules, and practices are required to produce meaningful and coherent democratic outcomes. In other words, if the impossibility theorems were indeed irrelevant and an unambiguous "best" social choice always existed, then a richer theory of institutions and collective choice would not be needed. The theory we present in Chapters 5 and 6 acknowledges and leverages this logical step as it presents an explicit social choice–theoretic notion of legitimacy.

Similarly, counter to the views of Pildes and Anderson, we argue that social choice theory is particularly well suited to studying value pluralism and incommensurability.[47] Value pluralism is, after all, captured by the assumption and utilization of the presumption of unrestricted domain. Particularly, when one notices that the impossibility theorems are about aggregation in the abstract and does not rely in any fashion on a linkage with individual or social "preferences," the imposition of IIA is *exactly* an acknowledgement of incommensurability.[48] In other words, Arrow's framework *embraces* the premise that the criteria to be aggregated are incommensurable, regardless of the nature of the criteria themselves.

In sum, Pildes and Anderson's arguments mirror those discussed earlier in focusing too narrowly on one particular interpretation of the impossibility results. Furthermore, arguments about the impossibility theorems that rely on the presumption that inputs to the aggregation problem need be individual preferences are fatally flawed. Of course, such a presumption is not inconsistent with the results, but it is logically incorrect to ascribe or deny substantive importance to the results themselves on the basis of such an interpretation.

On the other hand, our interpretation – which we develop in the following chapter – is wholly consistent with many of the arguments forwarded by Pildes and Anderson. Indeed, we believe that our approach productively formalizes and enhances many of their arguments. Similar to the match between our arguments and those of Riker, our approach comports with that of Pildes and Anderson in structure and starting points but yields entirely opposing conclusions. As noted earlier, we will argue in the following chapter that Arrow's independence of irrelevant alternatives condition captures precisely the kind of value incommensurability with which Pildes and Anderson are preoccupied. Moreover, the theory of legitimate choice we present in the second half of the book focuses exclusively on the linkage between collective choices and the reasons supporting those choices and does so within a purely social choice–theoretic framework.

[46] Pildes and Anderson (1990), p. 2200.
[47] *Ibid*, pp. 2143–2166.
[48] We return to this point in more detail in Section 3.3.

2.5 RIKER AND HIS CRITICS: UNLIKELY ALLIES?

In the end, we believe that Riker and many of his critics ask very little – indeed, too little – of democracy. To Riker, democracy is a "second-best" system in which collective choices do not necessarily reflect anything meaningful about the values of the societies they govern. Democratic procedures simply let people throw the *really* bad politicians out of office. While democratic institutions and procedures may enable collective decision making, they can confer no legitimacy to the resulting decisions themselves. To Mackie, democratic procedures are, in a very real sense, irrelevant. Because the preferences of individuals in society are structured so similarly in terms of either what people explicitly want or at least in terms of how individuals perceive the issues under consideration, there always exists a best outcome, and all we ask of democracy is to find it. Mackie argues that many reasonable systems are capable of finding such an outcome – in which case, problem solved; democracy is easy. The deliberative democrats argue in a similar vein that the process of deliberation induces preference structuration and thus returns us to Mackie's world in which there is a best outcome.

We take a very different view than Mackie, Riker, and the deliberative democrats by arguing that it is the possibility of cycles that makes democracy *meaningful*. The fact that there may be no best decision, that any choice we make can always be deemed inferior to other possible choices on the basis of criteria that society collectively deems relevant and important, is precisely what makes democratic decision making challenging and significant. Our argument bears some similarity to Anthony McGann's argument that Riker's critics yield too much, that all of these individuals presuppose that the widespread existence of cycling undermines democracy.[49] However, McGann argues that cycling *strengthens* democracy because the existence of cyclic preferences coincides with the existence of multiple alternative majorities that can each win and lose on different issue dimensions; deliberation and compromise will ensue as winners acknowledge that they must compensate losers sufficiently so as not to be quickly undermined themselves.[50] McGann argues that institutions limiting cycling, and majority rule cycling in particular should be viewed as suspect because they necessarily advantage certain individuals and alternatives over others. Majority rule, as a principle, is the unique decision process that is procedurally fair. The downside of majority rule, McGann acknowledges, is that it fails to produce a unique outcome.

While we find McGann's argument reasonable and persuasive in many respects, one question remains: How can an outcome – one that was selected via majority rule for example – be justified as a legitimate social choice when other outcomes clearly dominate it via the same majority principle? Here McGann is

[49] McGann (2006), p. 74.
[50] Miller (1983) and cited in Mackie (2006), p. 9.

less specific and says that the set of outcomes attainable via majority rule can be narrowed to the uncovered set, and that particular choices "... will result in part from bargaining."[51] While we agree that the impossibility theorems tell us that one should not hold out hope of a single, best social choice, our response to this is more in line with that of Pildes and Anderson: the possibility of cycles necessitates an approach to collective decision making in which reasons, or explanations, play a central role. Intransitivities arise when for any decision A, there is some other decision B that dominates it. Accordingly, justifying (or legitimating) the selection of A requires the provision of a rationale not only for the selection of A but also for the failure to choose B. In Part II of this book, we present a theory in which reasons or explanations along these lines are treated as a fundamental aspect of legitimate democratic governance.

To our knowledge, *Democracy Defended* represents the most comprehensive critique of the "Rikerian" interpretation of social choice and provides, along the way, particularly damning indictments of each of Arrow's axioms. These indictments synthesize decades' worth of arguments made by numerous scholars against the real-world relevance of the social choice enterprise. By nature of its scope and the tone and the pointedness of its attacks, we feel that the ball is now in our court, and we dedicate the following chapters to our aim of defending Arrow's axioms against the charges leveled at them and to show, counter to Riker's claim, that the impossibility results are wholly consistent with a meaningful conception of democratic choice. We do this through the best means we know how: through the use of social choice–theoretic concepts, tools, and arguments. Ultimately, our interpretation of the impossibility theorems is intended to motivate our own theory of democratic decision making that follows in the second part of the book.

[51] McGann (2006), p. 210.