ABSTRACT

Title of dissertation: PUBLIC OPINION,
POLITICAL REPRESENTATION,
AND DEMOCRATIC CHOICE
Michael J. Zenz, Doctor of Philosophy, 2015

Dissertation directed by: Assistant Professor Eric Pacuit
Professor Christopher W. Morris
Department of Philosophy

In this dissertation I argue that political representatives have duties to be responsive to public opinion in their policy decisions. The existence of this duty, I claim, is a basic requirement of a truly democratic system of government. In chapter 2, I show that several standard versions of democratic legitimacy require political representatives to "respect" public opinion. However, I argue that a particular version of political legitimacy, based upon popular sovereignty and the importance of self-governance, provides an especially useful background for understanding what this "respect" must mean. In chapter 3, I argue that respecting public opinion requires political representatives to integrate public opinion information into their policy decisions. According to one of the standard views of political representation, the liberal conception, representatives deciding between policy alternatives should balance what they believe to be in the interests of the public against public opinion. I argue that this is the only adequate theory of political representation. Although this view of political representation is often discussed in the literature, it is less

often given a mathematically precise form. Therefore, I present a formal model of such a balancing procedure, and this reveals several important formal requirements that a conception of public opinion must satisfy; most importantly, it must account for instability in the expression of public opinion, individual differences in opinion strength, and it must be representable along a cardinal scale. Standard measures of public opinion do not satisfy these requirements. I argue that if such a model of public opinion cannot be formulated, then the liberal conception of political representation is incoherent. In chapters 4 and 5, I present a model of public opinion based upon Thurstonian scaling techniques that fulfills the necessary formal requirements. Finally, in chapter 6, I discuss several important implications this model has for the measurement of public opinion, the use of public opinion by political representatives in policy deliberation, and other problems in social choice theory.

PUBLIC OPINION, POLITICAL REPRESENTATION, AND DEMOCRATIC CHOICE

by

Michael Joseph Zenz

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Assistant Professor Eric Pacuit, Chair/Co-Advisor
Professor Christopher W. Morris, Co-Advisor
Professor Peter Carruthers
Associate Professor Dan Moller
Professor Stanley Presser
Associate Professor Piotr Swistak

Dedication

For TLT.

Acknowledgments

Besides my advisors and the members of my committee, who all contributed to the success of my project, I would also like to thank Michael Morreau for introducing me to social choice theory and Rick Cai for introducing me to the psychophysics models that play a crucial role in this dissertation.

Table of Contents

List of Figures

Chapter 1: Introduction

Political representatives in modern democracies take many considerations into account when making policy decisions. Some are pragmatic, such as how a given legislative vote will be interpreted and evaluated by their constituents, most loyal voters, party leadership, and most likely campaign donors. These people greatly influence a representative's ability to be reelected and progress up the legislative hierarchy, and so it is reasonable to expect a representative to keep all of these individuals in mind when she is deciding upon which public policies to support. Other considerations arise from representatives' beliefs about what is best for their constituents, or what advances the common good. Finally, some considerations arise from their role as *representatives* who are responsive to the policy judgments of their constituents. Although all of these considerations are likely relevant for explaining the *behavior* of political representatives, and so are of great interest to political scientists, only a few are relevant to the representatives' *proper* role in a representative democracy. In other words, only some considerations that influence political representatives are required by democratic theory, while others may actually be excluded by it. In this dissertation I will explore the subset of the considerations that I think democratic theory requires political representatives to include in their

deliberations about what policy actions they should take. More specifically, I will argue that democratic theory requires that political representatives weigh public opinion against their estimations of what is in the common good when they make policy decisions. The bulk of this dissertation will be spent showing how such a model of political representation can be made coherent given a number of possibly devastating problems with such a conception of representation.

An examination of any major policy decision quickly reveals the diversity of considerations that representatives must often weigh in their decisions. One decision of particular interest in this regard was the 2010 passage of the health care reform legislation in the United States called the *Affordable Care Act* (ACA); I will use an example modeled after this piece of legislation throughout this dissertation, because I think it provides particularly salient examples of the weighing procedure between different considerations that I take to be central to the legislative actions of political representatives. A hint of this problem is easily apparent. In the spring of 2010, for instance, Congress was controlled by the Democratic Party after substantial election victories in the 2008 election. The health insurance provisions of the sort provided by the ACA had long been a goal of many members of the Democratic Party, and the legislation was seen by them as being in the common good. Passage also seemed to be supported by what many considered a significant electoral *mandate* that resulted from the 2008 national elections, which featured health care reforms similar to those in the ACA on the electoral platform of both the Democratic Party and its Presidential nominee Barack Obama. However, these considerations in favor of passage of the ACA were in conflict with some measures of *public opinion* at the

time. The ACA was very controversial, with opinion polls showing that a majority of Americans were against its passage, with the so called "individual mandate" that all individuals buy health insurance being particularly divisive.[1] Legislators faced these many conflicting considerations when deciding how to vote. Of particular interest to my project is the set of considerations bearing on their roles as legislators and representatives; situations such as the passage of the ACA make it unclear how political representatives should balance these conflicting considerations so that that they can best perform their role as representatives.

There is nothing unique about the decision process that representatives faced when developing the ACA; the question of how political representatives should deal with conflicting considerations is important for understanding how political representatives should act generally. In this dissertation I will examine the way in which political representatives should integrate different considerations into their policy decision process. In particular, I am interested in how representatives should integrate public opinion into their decisions and how this might be balanced against their estimates of what is in the common good. I will argue that political representatives are in a position of leadership, such that they should often use their own judgment about which policies will be for the best of the community at large, but they must also be receptive to the policy judgments of their constituents; I will refer to these judgments as the *public opinion*. In fact, I will argue that political representatives have a *duty* to integrate public opinion into their policy decisions. However,

[1]There was intensive public opinion polling during the passage of the ACA. For instance see the CNN Opinion Research poll released on March 22, 2010.

as I will show, any precise decision process that abides by those two principles must be capable of using measures of public good that have enough information content such that a *weighing* between the representative's estimates of the common good and public opinion is possible; it will turn out that this requires that public opinion be representable using *cardinal* (or numeric) values. Such a model of public opinion will face a number of difficult conceptual and technical challenges that will occupy much of the dissertation. For instance, a suitable balancing is only possible if public opinion is capable of being represented *cardinally*, something that many claim is impossible. I present a method by which public opinion can be represented cardinally, thus showing how political representatives can fulfill their duties to respect public opinion. By the end of the dissertation, I will be in a position to provide a coherent model of policy decision that is able to account for the integration of the diverse considerations political representatives face when making complex policy decisions.

In chapter 2, I show that a duty of responsiveness to public opinion is implied by a number of popular theories of democratic legitimacy, including the minimalist democratic tradition (Schumpeter, 1950; Dahl, 1991), deliberative democracy (Habermas, 1992; Cohen and Sabel, 1997; Pettit, 1997), and those theories derived from popular sovereignty. I then argue that a basis for democratic legitimacy in popular sovereignty, inspired by that of Rousseau, should be preferred. In democracies, because the people are sovereign, political representatives must respect public opinion or otherwise they violate the rights of members of the public to *govern themselves* — this right is a species of the right to personal autonomy. I show that the existence of a right of self-governance implies the existence of a corresponding duty

of some set of political officials to be responsive to public opinion; such officials are typically called *political representatives*. This theory also provides reasons for political representatives to use their best judgment about what is in the public interest, and not just public opinion, when making policy decisions. Therefore, the duty to respect public opinion requires political representatives to balance these two criteria when making decisions. This is similar to the "liberal" conception of democratic political representation, which originates from Pitkin's (1967) conceptual analysis of political representation in the western context. According to this view, a political representative must balance her "mandate" (public opinion about what should be done) with her "independence" (what she thinks is in the best interests of her constituents). I argue that representation so conceived allows political representatives to use their superior knowledge of policy issues and position within a small deliberative body to make decisions that better track the public good, while at the same time respecting the rights of the public to govern themselves.

I spend the rest of the dissertation responding to the worry that the liberal conception of political representation is incoherent. In chapter 3, I develop a formal model of policy choice that explicates this possible incoherence. This model shows that a mismatch between the *levels of measurement* of public opinion and measures of the common good is one important source of this supposed incoherence. When one *balances* one decision criterion against another, that balancing procedure must be sensitive to both differences in *weights* between those criteria as well as differences in *scale magnitudes* between the various alternatives. For instance, assuming that two criteria are weighed equally, if one criterion places some alternative x on a scale

well above another alternative y and the other criterion places y very slightly above x, then the resultant decision should somehow be sensitive to these differences (and likely rank x above y). Without sensitivity to differences in magnitudes, a system of weights in a decision procedure using two criteria does nothing more than choose which criteria is the *dictator*. This would force political representatives to choose to follow either their mandate or their independence but would never allow them to balance those decision criteria. The measure of what is in the common good in the model is almost certainly cardinal, given that most cost-benefit analyses (its real-world corollary) provide such cardinal measures. However, public opinion measures are typically only *ordinal*. I show this by examining three methods by which political representatives typically gain knowledge of public opinion: voting, direct communication, and public opinion surveys. All of these yield ordinal measures of public opinion, which are inadequate for inclusion in a balancing procedure.

In order to save the balancing conception of political representation, I must show how public opinion might be conceptualized and measured using cardinal measurement scales. This task is far from trivial, and because measurement scales are just special types of *utility* scales it requires me to show how interpersonal comparisons of cardinal utility might be made significant, thus producing an aggregate (or *social*) utility scale. This process would therefore produce a *social* utility scale representing public opinion from *individual* utility scales representing individual opinions. It is thought by many that the insignificance of such comparisons render the measurement of cardinal social utility impossible.[2] In chapter 4, I show how this

[2]This is the view of standard social choice theory after Arrow (1963).

problem can be avoided by carefully defining the source of individual utility scales as *attitudes* about public policies. I formulate an attitude model that explains the source of individual utility differences, and then show how those differences can be measured through series of pair-wise comparisons of alternatives. This serves as the basis for comparisons between individual utility scales. Because all individual attitude sets share similar underlying structures, there exist significant metrics (central tendency and dispersion of utility values) that can be used to normalize all individual utility scales. Therefore, at least within the context of attitude measurement, it is possible to make significant comparisons of cardinal utility. In chapter 5, I show that once a number of normative assumptions are made about which features of attitudes are *normatively* significant — and I argue that these assumptions are plausible within the democratic context — a political representative can formulate a cardinal utility scale representing public opinion. Such a scale could then be balanced against some measure of expected welfare in a social decision process.

This dissertation has implications for a number of research areas. I address these in chapter 6. The most direct implications are in democratic theory and public opinion survey methodology. I argue that a necessary feature of democracies is that some political officials involved in policy decisions respect public opinion. However, a method that is capable of capturing the full information content of the public opinion model that I present would likely require far more demanding measurement methods than are currently employed by survey methodologists, and likely more demanding than is feasible in real-life measurement settings. I present a number of simplification methods that might track the measurement of public opinion that I

present, though they may not completely capture the richness of the model.

I next address the implications of the model on deliberative democratic theory. Some might worry that any introduction of public opinion into legislative settings, in which deliberation about policy is central, will detract from the rationality of the policy formation process. After all, if political representatives are to come to a rational policy decision, one that is responsive to the best available arguments and reasons, then public opinion that has been separated from the reasons behind it may in fact make legislative decisions less rational. According to this view, there seems to be no space for public opinion in a deliberative institution. I argue that the use of public opinion can be integrated into decisions within deliberative institutions. I point out that within a society that features informal deliberative settings throughout, public opinion can in fact be seen as the result of a deliberative process, but one that has not been controlled to the degree as that found in a legislature; this should give public opinion some status in the public decision once deliberation has ended and a decision must be made. However, public opinion also gives representatives information about their constituents that might aid them in determining what arguments may ground public opinion. It will often be reasonable to include such arguments within the legislative deliberation process. Finally, communication between representatives and their constituents often resembles a deliberative setting, with representatives both informing the public with their arguments and being informed in return.

I then address an implication of the method of finding social utility on models in social choice theory, where cardinal utility representations are typically avoided as

measures of preference. It is generally thought that no significant way exists to make interpersonal comparisons of cardinal utility. I show that if more is known about what the utility is used to represent then it becomes possible to make significant interpersonal comparisons of utility. However, social choice theory is typically neutral to what underlying phenomenon utility represents. I provide a case in which abandonment of that neutrality may allow for a more powerful analysis of social utility using social choice theoretic tools. A similar abandonment of neutrality in social choice analysis has been suggested by Regenwetter et al. (2006), who argue that *actual* voting patterns take certain stereotypical forms such that they are not susceptible to several voting paradoxes. The method I use in this dissertation adds credence to the view that social choice analyses may often benefit from making certain assumptions about phenomenon being measured.

The central argument of this dissertation can be summarized as follows: Any representative democracy requires that all citizens be represented by some political official(s) who have a role in policy decisions. This fulfills the democratic requirement of self-governance. Because political representatives are also tasked with coming to decisions that are in the best interests of the public, and the public is sometimes misinformed about what policies are in their best interests, the proper *representation* of the people requires officials to balance what they believe is in the best interests of their constituents with the wishes of those constituents. A failure to adequately balance these constitutes a failure of political representation. However, a number of problems with the basic structure of *public opinion*, revealed primarily by work in social choice theory and survey methodology, makes such a balancing

procedure seemingly impossible. I present a model of public opinion that eliminates the possible problems with the balancing procedure. This includes a model of individual political attitudes, a model that represents individual attitudes in terms of utility scales over policy alternatives, and a social utility model that allows for the interpersonal comparison of those individual utility scales. I show that it is possible for public opinion to have a coherent role in the decision making of political representatives.

In the next chapter I will begin my argument by showing that respect for public opinion is crucial for any democratic theory that purports to allow for self-governance, something that I will argue is a crucial part of any adequate theory of democratic legitimacy.

Chapter 2: Democratic Governance in the Modern State

In this dissertation I will argue that political representatives within democracies possess duties to respect public opinion. Furthermore, 'respect' should be understood as a particular type of *responsiveness* to public opinion. In the *next* chapter I will give a formal model of the way in which representatives should respect public opinion. In this chapter I will argue that this particular way of respecting public opinion is crucial for justifying the legitimate authority of democratic institutions. There exist alternative ways of understanding respect for public opinion, but I will show that they do not provide this crucial role in explaining democratic legitimacy. Only a form of respect for public opinion that permits a degree of *self-governance* by the public could provide this *crucial* part of the justification for why democratic governments possess the political authority that they are commonly thought to possess.

There are many ways in which political authority might be legitimated. The existence of such a legitimate authority gives reasons to those who are governed to abide by (at least some of) the commands or laws given by that authority. In some social contexts the authority of, for instance, a even well-functioning dictatorship might be perfectly justified by its beneficial consequences to society. However, in

otherwise well-functioning political societies a dictatorship will not possess legitimate authority. In the social contexts in which most modern western states exist mere *beneficial consequences* do not seem sufficient to justify the authority of a state. Another species of justifications originates in the governing authority, or sovereignty, of the public itself. For instance, a democratic government might be said to have authority because it permits the people to govern themselves. The fact that one has been involved in the creation of a law or issuance of a command through a democratic process may give one reason to abide by those laws or commands. This is *democratic* legitimacy. One question I will attempt to answer is just what sorts of institutions are required in a government that is in part rendered legitimate by the form of its democratic institutions. I will argue that, at least in a state with large a population, such a government must have representative institutions in which public opinion is respected during the formulation of laws and commands. Furthermore, if this "respect" is to permit self-governance, it must include public opinion in the policy decision procedure.

2.1 Political representation and respect for public opinion

It will follow from the requirement of self-governance and the role of political representatives within representative democracies that any political official who serves the role of a *political representative* has a duty to take the public opinion of her constituents into account when making policy decisions. This duty is distinct from any duty she might have to do what (she thinks) is in the best interests

of those constituents or in the common good. Additionally, it addresses *individual* rather than institutional prescriptions. I will assume that political officials are working within a western democracy that is reasonably just and democratic (e.g. Britain, The United States, Germany, France, Japan, etc.). However, I will work on a more abstract level than concerns of how the opinion and interests of individual constituencies might relate to the common good of all of society, or how the duties of different types of political officials may vary within and between different political institutions. The purest case of a representative political official I have in mind is an elected "law-giver" who "answers to" all of society, though no such official exists in any actual modern democratic system. A more concrete case may be an "At-large" city council member or other member of a legislature who is elected by the entire public. But I think my conclusions will be general enough such that they can apply to more complex cases, such as members of a legislature who each represent different constituencies. Though such examples bring in complications, such as how different constituencies come to an agreement with one another, they will not substantially detract from my central claims.

A system of government cannot merely operate with concern for the *interests* or common good of the public for it to satisfy the requirements of possessing democratic legitimacy. Plenty of non-democratic regimes purport to, and sometimes actually do, aim at the common good. It is conceivable that in obviously non-democratic regimes, in which the people have no say in who governs or what policies are instituted, public opinion may actually play some role in public decisions as a measure of the *interests* of the public. But this does not make such a regime

democratic, because the use of public opinion in this way is simply incidental and could be eliminated at any time if some other better measure of the public interests were discovered. Similarly, some regimes that feature elections of officials may permit little role for public opinion in decisions.[1] Regimes in which elected officials make decisions with little concern about what the public actually thinks also seem importantly undemocratic. It is this last case that will be of particular interest to this project. I will have to give good reasons to think that free and fair elections do not provide a sufficient basis for democratic legitimacy.

A representative system of government in which officials wield substantial power to govern in accordance with the common good but also requires them to consider public opinion in their decisions has been called *"liberal" representation.*[2] My argument about the proper interpretation of self-governance in the legislative context relies heavily upon this conception of political representation, which requires that representatives balance the views of their constituents about what to do against their own views about what is in their constituents' best interests. I will defend this view here but I must also show that a number of popular alternatives to it are not adequate; there are many ways in which representatives can be thought of as "respecting" public opinion, and many of them do not require the consideration of public opinion that I will advocate here. As I will claim, they also do not require that the people are able to govern themselves in order for their opinions to

[1]Note that my claims here are about conceptual issues and not about how elected officials actually behave or what would maximize their prospects of reelection in the form of Mayhew's (1974) analysis of rational congressional behavior.

[2]This usage was made standard in modern political philosophy by Pitkin (1967), and is derived from its origins in Madison's writings – for instance in the *Federalist Papers*, especially #51- #58, #62, and #63 .

be "respected."

"Respect" for opinions goes beyond respect for the constituents *themselves* as persons or citizens. Even a dictatorship could show respect in this way. I will argue that 'respect' must refer to respect for particular *features* of the constituents, specifically their policy opinions. I think the following interpretations of giving "respect" to the opinions of the people are the main competitors, with the first being the interpretation that I will advocate:

Self-Governance The actual opinions of the public play a role in policy decisions.

Minimalist The government (and/or individual officials) voluntarily relinquishes power upon losing a free and fair election.[3]

Deliberative The reasons and arguments the people have in favor of their opinions are included in the deliberative procedure by political officials.

I will argue that the last two interpretations of 'respect' for the policy opinions of constituents are less satisfactory than the first. However, I will also show how any of these interpretations require the adequate measurement of public opinion in order to be adequately democratic. Therefore, what I do in this dissertation will be relevant to proponents of all three interpretations. As I proceed it is also important to realize that I am not refuting the use of any conception of democracy as a solution to various problems of institutional design, but only in so far as they are

[3]This general view is pervasive in political science and political theory. For instance, see Schumpeter (1950); Riker (1982); Dahl (1991). Also see Przeworski (1999), who uses the term to describe this type of democratic theory.

used as a fully adequate source of democratic legitimacy. The *Minimalist* conception of democracy provides good reason for frequent elections and the *Deliberative* framework provides reason to require that legislation be discussed in deliberative bodies, but neither framework adequately explains how democratic institutions permit the people to govern themselves. I will show that self-governance is central to democratic legitimacy.

2.1.1 Self-governance

When a system of government does not allow the people being ruled by it to govern themselves it detracts from the legitimacy of that government's authority to rule.[4] This is a claim I find very plausible, but it requires some explication to know just how the political authority of a government might arise from a representation scheme that permits self-governance. The explanation I will give here gains inspiration from Rousseau's theory of popular sovereignty, though I will rely heavily upon a particular contemporary interpretation of this well-known theory.

In *On the Social Contract*, Rousseau claims that a government derives legitimate powers to govern from the sovereignty of the individuals that compose the

[4]I do not take a stand here on whether any political theory can show democratic authority to be *legitimate* in the strong sense that contract theorists have often conceived of it, or even whether democratic states possess the sort of authority that they frequently claim to possess. According to this strong sense of legitimacy, if a state is legitimate then those individuals subject to its laws and commands have reasons to abide by them in every instance in which they apply. There is some reason to doubt that these standard forms of legitimacy and authority are often possessed by states, or even could possibly be possessed by any real world state. One major problem, for instance, is that many forms of legitimacy seem to rely upon the express consent of the governed, something most citizens of states rarely give, and theories that only rely upon tacit consent have a difficulty showing how merely tacit consent is sufficient to produce political obligations in the subjects. See Morris' (2002) criticism of legitimacy's status as the foundation of political authority and Simmons' (2001) criticism of the existence of political authority in most states.

community over which it governs. Rousseau thinks that this has the consequence of allowing individuals to be free even when they are bound to follow the laws and directives of the government. The sovereignty and authority of the state are derived from the collective sovereignty of the people and this is the way in which the people are said to govern themselves. Rousseau might have interpreted self-governance as lacking policy content (or direction), such that the people govern themselves insofar as they consent to be governed. This then might be similar in structure, though certainly not in detail, to other social contract views like that of Locke. However, Rousseau — at least according to the interpretation that I will introduce shortly — imagined that self-governance also provided *content* to governance in addition to consent. He claims that the wills of the public themselves play a role in providing a foundation for the political authority of the government. However, the connection between the wills of the citizens and policy decisions by the government is far from obvious. The primary guidance Rousseau provides in this regard is that legitimate actions of a government follow the *general will* of the public. The precise interpretation of the *general will* is an issue of great controversy for Rousseau scholars, and although it is not my intent to wade too deeply into it, a reasonable interpretation will be important for my use of the general will as a model of self-governance.

Gopal Sreenivasan (2000) provides a particularly insightful interpretation of the general will that also gives a very clear picture of just how the general will might be linked to public opinion. According to Sreenivasan, the general will is the decision that would arise from the public involved in a structured deliberation. One crucial passage for this interpretation comes from *On the Social Contract*, Book II,

Chapter 3:

> There is often a great deal of difference between the *will of all* and the *general will*. The latter considers only the general interest, while the former considers the private interest and is merely the sum of private wills. But remove from these same wills the pluses and minuses that cancel each other out, and what remains as the sum of the differences is the *general will*. If when a sufficiently informed populace deliberates, the citizens were to have no communication among themselves, the general will would always result from the large number of small differences, and the deliberation would always be good.[5]

An important distinction here is between the *will of all* and the *general will*. Each individual has a particular will, or set of wishes and preferences, for how the government should act, many of which are self-interested. For instance, if a city is deciding upon a location for a new landfill, those who live on the west side of town will prefer that it be located on the east side, and those who live on the east side will prefer that it be located on the west side. The aggregate of such individual wills is the "will of all." But this is not the "general will" that Rousseau thinks is the expression of popular sovereignty. In the above passage there appears to be two expressions of what the *general will* might be. The first, the "sum of the differences" that remains when one takes away from the "same will the pluses and minuses that cancel each other out", is rather cryptic. And in any case it

[5]pp. 31-32. Emphasis is my own.

will turn out to be less informative for determining what sort of institutions might instantiate popular sovereignty and self-governance than the second formulation. Sreenivasan concentrates on this second formulation of the *general will*, which describes the decision process that instantiates it. This is a deliberative process in which individual opinions are arrived at independently (without any "communication among themselves"), and in which all individuals within the deliberation are "sufficiently informed." Sreenivasan provides a more complete list of features of a deliberative setting that produces the *general will* from a wider reading of *On the Social Contract*, but does not significantly modify the picture provided in the above passage:[6]

1. The subject matter of the deliberation is perfectly general.

2. The conclusions of the deliberation apply to all the members of the community.

3. All the members of the community participate in the deliberation.

4. All parties to the deliberation think for themselves.

These constraints on the deliberative setting are not likely to shock anyone familiar with the deliberative democratic literature, some of which I discuss in section 2.1.3. Similar constraints characterize most good deliberative settings that are meant to determine what policies are in the common good. Rousseau himself seems to have thought that the general will would *always* track the common good, but this brings about a well-known tension in his view. It is unclear how *any* deliberation by

[6]Sreenivasan (2000), pg. 574.

actual citizens, even one that is extremely well structured to facilitate the influence of the best arguments, could *necessitate* the adoption of the policy that matches the common good of the public. People often have false beliefs about what is in their own interests let alone the common good, even when they are perfectly warranted to have their beliefs. The fact that the "common good" is often ill-defined and nebulous makes it all the more difficult for a deliberation to be directed at the same conception of it. It seems plausible that individuals taking part in the deliberation could reasonably have different conceptions of the common good. Therefore, unless one adopts a vacuous view of the common good, one that simply equates it to the content of the *general will*, one is forced to think that Rousseau's view faces a serious problem. The content of the *general will* cannot both be determined by the result of an actual deliberation and also always directed at the common good.

Although the tension between the *general will* and the common good is a well known problem for Rousseau's view, it is not a problem for my own use of the *general will*. There are a number of reasons to think that the *general will* so conceived will *tend toward* the common good of the public, and likely better than any other decision process; this is enough for my purposes. The tension in Rousseau's view stems from his insistence on an identity relationship (or something similar) between the content of the general will and the common good. But if this is relaxed to a relationship of *correlation*, then the conceptual problem with the general will disappears. What is left is an empirical concern about how strongly the general will correlates with the common good, and how strong this relationship must be in order to justify the reliance on the general will in public decision making. If this is only

a weak correlation then there is a reason to avoid the use of the *general will* in the formulation of public policy. But if it is well correlated with the common good then the *general will* seems to both have the benefit of tracking the common good closely in addition to permitting self-governance.

Up until this point I have not explained the concept of the "common good," A clearer conception of the common good will be important to determine whether the *general will* as I have interpreted it here tends to track it. I have already said that one important feature is that it is not simply an aggregate of individual wants or preferences. Such an aggregation is the *will of all*, which Rousseau explicitly states is distinct from the *general will*. Given that the *will of all* can reflect private rather than public interests this would be a poor proxy for the "*common* good." There are many alternative versions of the common good, and I do not think success of this project relies upon settling upon a single version, but the consideration of two main conceptions will be helpful to determine whether the *general will* tends to track it. The common good may be in those interests *shared* between individuals, and in this way arise from the individual interests themselves. For instance, individuals in a community may share an interest in a strong economy and clean air, though not in the prosperity of a majority at the expense of the prosperity of the minority.[7] This will be the way in which I will conceive of the common good. However, there might also be goods for *communities*, that may not arise from the interests of any individuals when they are not taken as members of that community.[8] For instance,

[7]This I take to be the conception of the common good favored by political liberals in the pattern of John Rawls.

[8]This is a concept of the common good which is similar to that advocated by communitarians. Will Kymlicka has a concise presentation of this view in his *Contemporary Political Philosophy*,

if the existence of a community or its way of life is threatened, obtaining what is in the common good of that community may require that its members sacrifice some of their interests to assure that the community survives. The first type of goods can be reduced to some set of individual goods, whereas the second type has an irreducible character.

There is reason to be optimistic that the general will is tightly related to the common good of either general type that I describe above. This optimism is based in two commonly held beliefs in political theory. Firstly, the policy preferences of the public are often the best source of information about what is in the best *interests* of the public. This is primarily relevant to the first *reducible* conception of the common good. Secondly, deliberative decisions are often better able than preference aggregation for arriving at decisions that are in the *common* rather than individual interest. This is relevant to both conceptions of the common good.

Robert Dahl argues for the acceptance of a principle that is relevant for how the general will might track the first type of common good; he calls this the "*Presumption of Personal Autonomy.*" This presumption states that "In the absence of a compelling showing to the contrary everyone should be assumed to be the best judge of his or her own good or interests" (Dahl, 1991, pg. 100). One might also take a more Millian view, that personal autonomy is not just instrumental to determining individual interests, but *constitutive* of an individual's interests.[9] Some interpretation of this principle seems correct, and I doubt many democrats would

pg. 220.

[9]John Stuart Mill expresses this in *On Liberty*, chapter 3.

doubt its truth, though some may claim that there are many cases in which there exists a "compelling showing" indicating that the public is not the best judge of their own interests, especially in, for instance, economic policy making.[10] Examples of where *public opinion* fails to track the common good often, it seems, amount to claims that public opinion does not meet the standards of the *general will*. Drawing from Rousseau, if the public fails to be "sufficiently informed" about some issue then that public opinion does not meet the standards of the *general will*. A similar failure to meet the standards of the general will arises if significant factioning develops such that individuals fail to make their judgments independently. In fact, it is unlikely that public opinion of a large society will ever come close to meeting these standards. As I will describe soon, it is not my intention to show that *public opinion* itself is an instantiation of the general will, but rather that it plays a role in forming it.

The second aspect of the general will that aids in its correlation to the common good is its *deliberative* nature. This is relevant to both types of common good that I mention above. According to Sreenivasan, Rousseau thinks that the purpose of the deliberative procedure is to bring out the "true wants" of the people rather than the wants that they may happen to espouse at any given time. More precisely, these are "...wants that the person has ratified through his own reasoning and that also correspond to his critical interests" (pg. 551). Given that I do not want to delve too deeply into Rousseau's particular conception of 'interests,' the concept

[10]For example, see Caplan (2008) for such an argument against the ability of the public to arrive at decisions about economic policy in the common good.

of "critical interests" might be interpreted as roughly equivalent to an individual's *actual* interests. It is, therefore, important to remember that individuals do not express their sovereignty through *any* preference they have, no matter how poorly informed it may be. A democratic theory based upon the general will is more restrictive of the preferences used in political decisions than a more naive form of populism that requires governments to act according to any crude wants of the public. In particular, the arguments and reasons provided in deliberation will tend to focus the deliberators to those issues and policies of shared interest, relevant to the community taken as a whole. For arguments to be accepted within a deliberation they must be acceptable to all who take part in the deliberation. With this in mind, a deliberative procedure can be thought of as focusing the attention of the public to that which is shared in common, and therefore the common good.

The explanation above only addresses how the tension between the common good and the general will (as interpreted as a structured deliberation by all citizens) might be resolved; one should be careful to see that this does not simply reduce the general will to a *tool* that tracks the common good. There is reason to care about the form of the general will apart from its ability to track the common good. One might possibly interpret Rousseau as formulating, in the general will, the proper goal for the government to pursue. Such a role for the general will need not be the explanation for how democracies have the authority to govern. If the sole reason to abide by the general will is the fact that it is well correlated with the common good, then a dictator using some sophisticated decision procedure might do this just as well as a popular deliberative body might. The *general will*, however, can only be

instantiated by some particular democratic process, and this in itself differentiates itself from the common good. The common good is a particular outcome rather than the result of some procedure. What makes the *general will* an instantiation of popular sovereignty is the fact that it permits the people to govern themselves: it promotes self-governance while at the same time directing the public to govern towards the common good.

I have not yet mentioned what is likely the most problematic aspect of Rousseau's formulation of the general will. If the general will is taken to depict a real (or even model) decision process then it seems to require that an entire political community deliberate on the issues in a single deliberative setting. This is one of the reasons why Rousseau suggests that small societies are best. In fact, Rousseau seems to explicitly exclude the appointment of political representatives to permit a deliberative setting of a reasonable size, stating that:

> ...since sovereignty is merely the exercise of the general will, it can never be alienated, and that the sovereign, which is only a collective being, cannot be represented by anything but itself. [11]

He clearly views an elected body of representatives as inadequate for permitting something like self-governance. Referring later in *On the Social Contract* to the English parliamentary system of the time, Rousseau states that:

> The English people believes itself to be free. It is greatly mistaken; it is free only during the election of the members of Parliament. Once they

[11] Rousseau, *On the Social Contract*, pg. 29.

are elected, the populace is enslaved; it is nothing.[12]

Rousseau, I think, is basically correct here. If the primary means by which the people exert influence over policy decisions is through elections of representatives then there is something seriously lacking in their ability to govern themselves. One could take Rousseau's criticism of representation to have striking negative consequence for Rousseau's view of representative democracies. If legitimate governments were required to govern over only small societies then few governments are legitimate. Perhaps the small 18$^{\text{th}}$ century city-state of Geneva, which Rousseau so admired, might qualify as legitimate but almost no modern states would. The natural remedy for this problem is an adequate view of political *representation* that shows how and to what extent the *general will* might be instantiated within a decision procedure within a representative governing body. The hope is that such a project can show how political officials and their deliberations might represent something like the *general will* of the public within representative legislatures. If the essential elements of Rousseau's system of legitimization are preserved within such a process, then it would seem that large representative democracies might also be in part rendered legitimate by way of popular sovereignty. This is explicitly contrary to Rousseau's view that wills cannot be represented.[13] I have doubts about many details of Rousseau's view about sovereignty, authority, and legitimacy, but

[12]Ibid. pg. 74.

[13]In *On the Social Contract*, Rousseau *does* seem to allow for political representation to be involved in the governance of the people. These representatives are then entrusted with acting in accordance with the general will. Note however that this form of representation does not allow for the will(s) of the represented to be present in the acts of the representative. Therefore, the *general will* cannot be represented, though representatives can govern in accordance with it. Therefore, the *government* may be composed to political representatives, though the *sovereign* cannot.

these will not play a significant role in my discussion of self-governance and political representation.

I will use the general will as a model for a democratic decision procedure that permits self-governance and tends to track the common good. A representative democracy based upon the above interpretation of Rousseau's *general will* is similar in many ways to the deliberative democratic framework of political legitimization. I will argue that these are compatible but that Rousseau's concept of popular sovereignty overcomes some concerns that one might have about the potential of deliberative democratic systems for being *undemocratic*. I will argue that political representatives in democracies have duties to respect public opinion that arises from the requirement that democracies permit the public to govern itself. I will also argue that Representatives can only be thought to "respect" the opinions of the public if they take public opinion into account when making decisions within a deliberative setting. Although I do not think that Rousseau would consider this use of public opinion a true instantiation of the *general will*, because it still does not allow the people to be directly present in the decision process, I think that it tracks the concept of the *general will* well enough to explain how it permits self-governance.

In the beginning of this section I stated that the concept of the *general will* would aid in the understanding of what it must mean for political representatives to respect public opinion by facilitating the self-governance of their constituents. I think some features of that decision procedure can already be made clearer. Any group of individuals can be thought of as having an *aggregate* will if individual wills are conceptualized as utility functions over the set of possible group actions (or

policies).[14] However, individual wills often conflict. In every society some public decisions will benefit some individuals more than they do others, or worse yet, benefit some at the expensive of others. Returning to my example of choosing the location for a landfill, someone who owns land on the east side of a city would rather a landfill be placed on the west side, whereas someone who owns land on the west side would rather it be placed on the east side. If the landfill is located on the east side, those on the west side will certainly benefit, whereas those on the east side may actually have been better off had no landfill been built at all. These two individuals have different self-interested preferences for where the landfill is placed, and the aggregation of all such preferences within a public can be thought of as the *will of all*.

If the will of all were used for public decisions, then it would result in something like a pure aggregation of votes; if those on the east side have one more vote than those on the west side then the public decision must be to build the landfill on the west side, even if there are good public reasons (perhaps the east-side location is larger) to do differently. But this mode of thinking is not the sort that public decision making should consist in; instead, individuals should restrict considerations used in their policy decisions to those relevant to what is in the best interest or common good of the *community, people, or state* as a whole. Accordingly, the *general will* requires that there is deliberation about the policy such that the decision tracks

[14] A utility function is just a preference relation over some set of alternatives. It is a technical concept, without any of the theoretical content that the term 'utility' often connotes in moral and political philosophy. Assume, for the moment, that there are no social choice theoretic problems with such an aggregate will. I deal with these issues directly in this dissertation, and will show that they can be avoided if the aggregate will is conceptualized in the appropriate way.

the common good rather than merely the private interests of the majority. When this is successful then the public wishes that are included in deliberation are only those that are relevant to public decisions that aim at the common good. However, notice that these public wishes, even when they are directed at the common good, will often fall short of the general will because they lack a substantive deliberative component and may contain serious rational deficits. Even if all the members of the public are only concerned with the common good, within large societies it is impossible for all citizens to rationally deliberate with each other. Therefore, the individuals who deliberate will not be coextensive with the public. This threatens to make the concept of the general will irrelevant to large modern democracies.

In my view, political representatives who are responsive in the appropriate ways to public opinion can instantiate something similar to the *general will*; I will argue that representative legislatures allow for a public to implement policies that reflect the opinions and interests of the people in a way that tracks the common good. This instantiates something like the general will and thus permits the public to express its popular sovereignty even when most members have no direct involvement in the deliberation (among representatives) that brings about public decisions. In the next chapter, I will describe a model of political decision making that I take to be an instantiation of the decision procedure Sreenivasan describes as bringing about decisions consistent with the general will that track the common good. I prefer this Rousseauian-inspired picture of political representation, but this is not a work of Rousseau scholarship and therefore I take the view to be attractive on its own regardless of whether it is an accurate interpretation of Rousseau's view.

Next I will argue that taking 'respect' for public opinion to require the facilitation of self-governance should be preferred over some other likely contenders. I will show that notions of respect from both *Minimalist Democrats* and *Deliberative Democrats* are inadequate. However, even if one accepts one of these other notions of respect, I think that the model of democratic decision making that I will present in this dissertation is relevant to how political representatives should act.

2.1.2 Minimalist Democracy

According to what I will call "Minimalists Democrats," democracies are better than alternative methods of governance, because they typically lead to more beneficial outcomes for the governed. Specifically, they are more likely to lead to stable societies that tend to track the interests of most people in society. This view is especially popular among political scientists. "Democracy" in this context is generally defined as the existence of regular competitive elections, nearly universal suffrage, and the peaceful transition of power in accordance with the results of those elections. But this view requires no substantive connections between the policy opinions of the public and the actions of the government. There is certainly a significant insight into how societies function in this tradition, but I will show how this cannot be an adequate explanation for why democracies might have legitimate authority when other forms of government do not. Such *Minimalist* democratic theories often assume, mistakenly as I will show, that more demanding forms of democracy are impossible or incoherent. Showing that this is mistaken will be the purpose of much

of this dissertation.

It should first be observed that *Minimalist* conceptions of democracy are typically based upon two theoretical foundations: a consequentialism about the purposes of government and a skepticism about the existence of a popular will and/or the ability of elections to convey it. The sorts of reasons minimalist democrats give for why democracy is desirable typically rely upon its supposed good consequences for the lives of the governed. For instance, Przeworski (1999) claims that generally only democratic states are capable of providing peaceful exchanges of power, and that such exchanges are highly beneficial to members of the state. Power exchanges dictated by democratic processes avoid the violence and resultant disorder common to non-democratic regimes changes. Because changes of power are inevitable within any society, a system that can eliminate violence and subsequent social disorder greatly improves the well-being of individuals living in the state. This notion of well-being can be defined in many ways, based upon a measure similar to that of utility in economics or a more complex notion of well-being. It can also be found through an aggregation procedure or a function sensitive to the distribution of overall well-being. However, whatever the details the underlying normative justification for democratic systems remain broadly consequentialist in character. Importantly, it is also contingent upon the physical and social environment in which the state exists. It is plausible that in some environments democracies would have few if any redeeming qualities. There is nothing good in itself about the democratic form of government.

A second central feature of minimalist conceptions of democracy is a skepticism

toward the ability of voting behaviors to convey information about the policy views and interests of the people. Schumpeter (1950) and Riker (1982) present two classic arguments for this. Schumpeter claims that the public generally has incoherent and uninformed views about most public policy issues and Riker argues that findings from social choice theory demonstrate that ordinal voting theory could not possibly convey coherent views from the public (even if those views existed) to political officials. This skepticism is one important reason why one might turn to more minimal justifications of democracy that do not rely upon the connection between the opinions of the people and the actions of the government. As I will argue in section 3.3.1, I share the skepticism with minimalist democratic theories that voting behavior can adequately convey the policy views of the public to political officials. However, I do not think that this limitation of *voting* should doom democratic theories that rely upon some connection between the actions of the government and the policy opinions of the public. I will claim that the problem with voting is primarily its inability to properly measure the policy attitudes (the concept I will give public "opinions") of the people, but that this is not reason to think that those attitudes do not exist in a form relevant to policy formation by political officials. Additionally, I will show in chapters 4 and 5 that public policy attitudes can be coherently conceptualized and measured.

If the skepticism about public opinion is dropped, then I think minimal democracy is no longer as attractive. What remains, I think, is the appeal of its consequentialist basis, at least for those inclined to such theories of political legitimacy. More precisely, there is a tendency for *Minimalist Democrats* to focus exclusively upon

consequentialist considerations when evaluating governmental forms. Although I agree that the consequences of a form of government should figure into its evaluation, I do not think that consequences exhaust the set of evaluative criteria. As I have explained above, I also think that the opportunity democracy gives to individuals to govern themselves is an important consideration in its favor. Furthermore, purely consequentialist justifications seem to support political systems that give very little control to most citizens. The psychological and social benefits of self-governance might even be produced so long as the people are made to *believe* they play a role in governing. A society might actually be ruled by a small group of ultra-competent civically-minded aristocrats. This elitist political system is *in principle* unacceptable in my view, but a minimal democrat would be forced to explain its unacceptable nature in terms of certain contingent facts: that it is inherently unstable or that a set of aristocrats will always (eventually) act in their own self-interest. It seems possible that an institution might be suitably designed to promote long-term stability and directedness toward the public good of such an aristocracy. In any case, there is considerable controversy about the relative harm to social well-being of elitist political systems.[15] Therefore, it may be difficult for the minimal democrat to condemn a political system exclusively for the system's consequences on the lives of those who are governed. Any adequate democratic theory should be able to criticize, for instance, something like Plato's vision of an ideal city-state (as my example above resembles) even if Plato is correct about how well a society so organized would thrive. I would say that such a system does not allow the people

[15]For a discussion of this see Peffley and Rohrschneider (2007).

to suitably govern themselves, but it would seem that the *Minimal Democrat* would have little to say against it.

Even if my criticisms of *Minimalist Democracy* are unsuccessful, if it can be shown that some coherent public will or opinion does exist in a form that could possibly guide policy decisions then political representatives should be responsive to that will. Robert Dahl's (1991) *Presumption of Personal Autonomy* is relevant here. He claims that one core justification for democracy is the immediacy and veracity with which every individual knows her own interests. Because each individual has better access to knowledge of her own interests than any other individual, public officials should generally accept each individual's assessment of what is in that individual's interests at face value.[16] I think this has intuitive appeal, certainly if the assumption simply amounts to the claim that individual assessments of interests should be used by political officials, in conjunction with other information, to make decisions. There are some contexts in which these assessments might simply be ignored, for instance in very technical matters about which the public is generally uninformed, but it calls for a presumption in favor of including this information. This would seem to support the use of public opinion by political officials if it can be shown that some coherent measure of public opinion exists in a form that can be measured. In this dissertation I will show that this is the case.

I have shown in this section that there is reason to deny the *Minimalist Democrat's* view of "respecting" public opinion. Furthermore, even if one accepts the

[16]Note that it need not be assumed that individuals have *special* (in the sense discussed by philosophers of mind) access to their own interests to accept this. It need only be assumed that individuals generally have better access to this knowledge than any other individual.

Minimalist Democratic view of "respect" there are still consequentialist reasons for the public opinion to be integrated into the decisions of some public officials.

2.1.3 Deliberative Democracy

The deliberative conception of democracy claims that the legitimacy of democratic decisions relies upon their origins in fair rational deliberative procedures, or the similarity of those decisions to what would be arrived at by an ideal deliberative procedure. This tradition is quite compelling, because it seems to show how a democracy might be both inclusive and rational. By permitting individuals equal access to the rational discussion that gives rise to public decisions, there is a sense in which the opinions of all citizens might be respected by such a deliberative process. This seems particularly true when all of the citizens of a social group are included in the deliberation, and when certain deliberative norms facilitate free, equal, and independent deliberation. Therefore, in its most idealized forms, deliberative democratic theory purports to show how the public can govern itself on a free and equal basis through a deliberative process. This resembles in certain ways the method of finding the *general will* that I have suggested above.

Deliberative democracy appears to match the sort of respect for public opinion that I take to be central to self-governance, at least in so far as those opinions are backed by reasons acceptable to all.[17] However, large complex societies that are organized into modern states do not permit the direct democratic deliberation

[17]Later I will show that my own conception of public opinion does not require that the reasons behind it be apparent, though the *entire* decision process by political officials should be rational in a way similar to what deliberative democrats suggest.

on the order of what Rousseau discusses, even at the local level (at least upon a superficial interpretation of his theory). Few public political organizations of any type could realistically operate as direct deliberative democracies and even fewer would satisfy the requirements of rough deliberative equality that the deliberative democratic model requires. For instance, even deliberation among all the citizens of a small town of 500 would be unlikely to satisfy the conditions required for that deliberation to point to *the general will*. And even one of the best known proponents of large-scale deliberations, James Fishkin, does not claim that public deliberations could ever meet something close to Rousseau's standards.[18] For instance, a Fishkin-style "deliberation day" only permits some *representative sample* of the population to deliberate about a set of issues. And those Fishkin-style deliberations that have been attempted do not typically meet the requirement that all individuals think independently; factions and deference to designated "experts" are prevalent. Although this sort of deliberation may be useful to the public as its members are forming opinions about various issues, it seems implausible that it could serve the role of a legitimate *legislative* body.

Even if *direct* democratic deliberation is impossible in most cases, it may be possible that deliberation among *political representatives* might approximate the requirements on deliberation. One way in which a deliberative process among political representatives might respect public opinion is if it gives equal consideration to the reasons and arguments for various policy positions held by the people. Additionally

[18]Fiskin has written much about his experiments with deliberation among actual citizens. For an introduction to this project see his 2009 book *When the People Speak*.

a rational deliberation process purports to give equal consideration to the *interests* of members of the public by being concerned only with the common good or public interest. I will consider several interpretations of deliberative democracy that I think are natural readings of deliberative theory compatible with political representation and I will examine how well they promote self-governance. Here I will explicitly consider the work of Jurgen Habermas (1992), Joshua Cohen (1989), and Philip Pettit (1997). Although these authors all have distinct views of deliberative democracy, different interpretations of their views can cause them to either diverge or converge. I think that these views, as well as most deliberative democratic views, converge to one of two distinct interpretations of deliberative democracy depending upon how they are explicated. Therefore, I will attempt to formulate different versions of deliberative democracy independently from the authors who purport to support them, though I will mention their views when they are relevant. I think that one interpretation is not adequately democratic and the other interpretation requires that representatives make use of public opinion in a way similar to what I will describe in this dissertation.

The first interpretation of deliberative democracy treats reasoned discourse as the primary mode of legitimating social decisions, a process that can be adequately carried out by political representatives. This interpretation arises most naturally out of Pettit's Republicanism, but I think it is also a natural consequence of Habermas' and Cohen's views if an attempt is made to make their idealized theories more concrete. Specifically, it arises from Cohen's (with Sabel, 1997) notion of deliberative "Polyarchy" and Habermas' discursive democracy when the deliberative ideal

is treated as a hypothetical condition that legislation must meet, rather than a description of the decision process. This interpretation of deliberative democracy is not sufficiently democratic.

The second interpretation arises from a treatment of the idealized condition of deliberative democratic theory treated as a model of the decision process; I specifically have in mind Habermas' *discursive* democratic model and Cohen's variant of this. However, I think that Pettit's Republicanism can also be interpreted in a similar way. Discursive politics requires that deliberation happen at two primary levels within society: in a formal setting among political officials and in an informal (and unconstrained) setting among all members of the public.[19] Although policy decisions are made by the formal process within political institutions, in order for such a system to be *democratic* there must be some means by which deliberation at the informal level affects deliberation (and consequently decisions) at the formal level. A reasonable explication of this process requires that *public opinion* be involved in institutional (and especially legislative) deliberations and decisions. Although all three authors mention that public opinion should play some role in legislative deliberation, they do not describe in significant detail how this actually should take place. Therefore, the theory of political representation and the duties of those representatives that I will provide in the next chapter might simply act as an explication

[19]Cohen and Fung (2004) place particular importance on deliberation among citizens rather than within a formal legislature, and their "Radical" democratic system places some formal constraints on the deliberation that takes place in the public. They imagine that deliberation among political officials within a legislature only has the purpose of providing suggestions and direction about the subject matter that should be discussed within the *local* deliberative forums open to all citizens. In many ways this instantiates the general will, and therefore does seem to allow for self-governance. However, I have doubts about how likely it is that such a system could adequately *govern* a large modern state.

of this view. In this way, I see the second interpretation as being closely related to a political system that requires self-governance through the use of public opinion by political officials.

The first interpretation of deliberative democracy places too much emphasis on legislative deliberation about *interests* at the costs of the actual policy *opinions* of the public. If committed to this interpretation, deliberative democrats would advance a substantive view of political representation that is undemocratic. Deliberative democrats often describe democracy such that there is no requirement that policy decisions are clearly linked to the wills, opinions, or attitudes of the people; they typically only require that people have some rational *recourse* against policies with which they disagree. However, without a link between the people and the method of policy choice itself, a political system cannot properly be called *democratic*. Rather, what is left is a *rational* social decision procedure with a stated end of advancing the common good or interest. But this ignores the value of self-governance, because it withholds from the people the right to have their considered policy opinions involved in the eventual decision.

Consider a political system in which the identities of those political representatives who present arguments as well as how they eventually vote are kept a secret from the public. There is some reason to think that such decisions would tend to be *more* rational, and better point to the public good, than those that are totally transparent to the public.[20] Deliberative transparency may, for instance,

[20]See Stasavage (2004) for a discussion of the benefits and costs of transparency in policy formation.

force deliberators to argue for policy positions that they know to be inadvisable just because some set of their constituents or political donors support those positions. It is not necessary to imagine a hypothetical case in order to see this in action; many bureaucratic bodies deliberate about which policies are in the common good, but are largely opaque to the public. And legislative deals between political parties are often struck in closed-door meetings. The deliberative democrat may claim that all deliberation should be public, because the people should be made aware of the reasons and arguments for the adoption of each policy. But this might be accommodated by imagining that a transcript is produced at each deliberative meeting, one which gives a complete account of the deliberation except with all names replaced by random numerical labels. There is little room for inclusion of public opinion in such deliberation, especially given that there is likely little public interest in some of the more mundane decisions made by bureaucracies, but as long as the deliberative process is guided by reasons and the public good this would seem to fulfill the deliberative ideal. However, all links between the public and the government are lost in this case, and members of the public are no longer able to effect change in the adoption of policies through the expression of their opinions and their votes, because they have no means of determining just how their representatives argue and vote. This would not even satisfy the conditions for *minimal democracy*. Such a system seems unsatisfactory from a democratic point of view. Of course, such a process may be perfectly adequate for administrative and bureaucratic decision making, but it is not adequate for a *legislative* process.

Pettit (1997) emphasizes an aspect of Republicanism that may be important

to the democratic credentials of deliberative democracy: individuals need to have avenues to *contest* public decisions that run contrary to their opinions on the issue.[21] Therefore, there should be some rational process in which public officials must defend their decisions to the public, as well as give public hearing to policy opinions of the public and the reasons that back them. In this way, members of the public might influence public decision making if they are able to provide arguments that are convincing to the legislative members. Although this opens the forum to the public, it subordinates the rational decision making of the public to that of the legislature. I think this is a crucial criticism of this type of deliberative democracy. The currency of deliberation is *argumentation*, and all members of the public are free to give arguments that will be treated equally. However, the value of arguments are judged from the top-down. In a system that is strongly deliberative rather than populist, decisions are made by legislators after they appraise the available arguments; when public opinion is largely ignored, then the legislators are permitted to be at the helm of an oligarchy with respect to the appraisal of arguments. I will argue that in order to avoid such an oligarchy the opinions of the public must be included into the deliberative process. When these opinions are the fair judgments of the public about what is in the common good or interest, they should be given standing in their own right.

The second (more populist) interpretation of deliberative democracy provides an opportunity to include reasoned public opinion in the public decision process. Habermas, for instance, discusses two "tracks" of deliberation within society: a for-

[21] Specifically, see pp. 183-200.

mally constrained type that takes place within political institutions and the informal type that takes place within the public. An important aspect of the theory is the flow of opinion between areas of formal and informal deliberation. Cohen (1999) has concerns about how effective such a flow might possibly be. Habermas himself seems to concentrate on the flow from the informal to the formal deliberative systems, especially in the case of important social movements. But Cohen thinks that this would limit, to an unreasonable extent, the degree to which the public could affect legislative deliberations. Similar in many ways to Fiskin's (2009) deliberative procedure, Cohen imagines a deliberative system he calls a "directly-deliberative polyarchy," in which many public decisions are made by deliberative bodies composed of members of the public, whereas general guidance and direction is given by a legislature. As I described above, I am skeptical of the capability of public deliberative bodies to make decisions in large societies, but I think that Cohen is correct that there must be more attention given to just how the results of informal deliberation in the public might flow up to formal deliberative bodies, and especially legislative ones. In the next chapter, I will propose one such a method by which this upward flow might occur. I think this is compatible with the second interpretation of deliberative democracy that I have described here, but is more easily captured by the view I described earlier in this chapter based upon the general will.

I will return to the role of deliberation in democracies in section 6.2.2, after I have presented my models of political representation and public opinion.

2.2 Respect for public opinion

What I have sought to do in this chapter is show that democratic political institutions should foster respect for the opinions of the public. I have prepared the way for my discussion, in the next chapter, of how exactly they should do this: roughly stated, political representatives should integrate public opinion into their deliberation and decision procedures. Although I think this is the case generally, this is especially important for *legislative* institutions to properly be considered democratic. Additionally, this respect should even extend to those opinions that are rationally *opaque* from the representative's perspective. Although most deliberative democrats would accept that arguments provided by members of the public should be respected, it is typically important that *reasons* and *arguments* are attached to public opinion when it flows up to the legislative process. I will claim that the very fact that the public comes to some opinion about some policies gives the legislative body reason to respect that opinion; by doing so, the legislature respects the people as sovereign in themselves and capable of coming to reasoned decisions without further rational assessment by the legislature.

All that I will attempt to show is that there should be a *presumption* that public opinion should be integrated into the decision process. Of course, there will sometimes be overriding considerations against using public opinion; for instance, if it advocates the violation of civil liberties or it is clear that the public is badly misinformed about a matter of great importance. And I am *not* advocating for a purely *populist* principle, whereby public opinion should be generally followed

without modification. As I described earlier in this chapter, I think that public officials play a vital role in using a deliberative process to decide upon policy that accords with the general will, and consequently aims at the common good. This will require a decision process that includes public opinion as an input, but is not fully exhausted by it.

One core consideration in favor of the presumption is based upon the existence of rational peer disagreement. To begin with, it should *not* be assumed that legislators are *significantly* more competent reasoners or have significantly better access to true information relevant to policy decisions than the public, at least not in the manner that adults are significantly better reasoners than children.[22] Such a large distinction in basic rational capabilities and access to relevant facts would undermine the very basis of popular sovereignty; when some small subset of the population is truly superior in some major respect to others in society then the grounds for equal rights to political participation become weaker. There have been some doubts as to whether most members of the public possess the basic sophistication required to make policy decisions.[23] And some have worried that such a disparity in competencies risks undermining the bases for the basic principle of popular governance in democracy (Friedman, 2006). But research into the supposed "wisdom of the crowd" should create skepticism about whether a single legislator could be a better judge of whether a policy advances the common good than her constituents

[22]I take this assumption from Dahl (1991), though I take it to be widely held.

[23]This is the classic view of Lippmann (1922), that was revitalized by empirical work following after Philip Converse's (1964) classic studies. Also see Converse and Markus (1979). Notice how this is distinct from Schumpeter's (1950) criticisms of democracy that focus on the *unwillingness* of the public to devote attention to questions of public policy.

as a *group*. For instance, large diverse groups often collectively have access to information and expertise that is unlikely to be possessed by a single individual.[24] Recent work in the study of public opinion surveys also provides some evidence for the view that the public, again taken as a group, is relatively rational when choosing between policies (Page and Shapiro, 1992). I have doubts about the general "wisdom" of groups when their judgments are unchecked by legislative deliberation, but I think that this research should create skepticism about the general presumption of an incompetent public.

If it is not clear that legislators are significantly more competent at making policy decisions than the public, or even individual members of the public, disagreements between legislators and the people should generally be taken as cases of genuine peer disagreement. Such a view gains even more credibility after a lengthy public deliberation about the policy issue has already taken place, and therefore it does not seem that continued group reasoning could alter the opinion of the public. Given that people bring different values and beliefs (as well as different priority orderings over them) to the decision process, it is possible that different individuals may arrive at different conclusions even if they all reason reasonably well. It is a matter of controversy in social epistemology how exactly peer-disagreement should be treated, but I think that it does give reason to back the *presumption* that officials should respect public opinion.[25]

[24]This has been most popularized by Surowiecki (2005). Though the basic idea is similar to Condorcet's well known "Jury Theorem," which proves that for a set of independent two-alternative forced-choice judgments, the average of the that set approaches the correct response as the size of the number of individuals making judgments (the size of the set) increases. See Lyon and Pacuit (2013) for a useful introduction to the "wisdom of the crowd."

[25]For instance, see Lackey (2008) for a discussion of this.

It is one matter whether *institutions* should be created that facilitate and promote the respect for public opinion, but quite another whether individual players within those institutions have duties to act in accordance with purposes of the institutions. Very few discussions substantively address the duties of individuals within democratic institutions that otherwise meet the requirements of democratic legitimacy (or are at least deemed democratically *adequate*). This is unfortunate, because most questions of proper governance occur within existing political institutions that are at least partially democratic (and presumably at least partially legitimate, justifiable, or adequate). Political officials must often decide how exactly they should act within those democratic institutions. The guidance that does exist is largely pragmatic in nature. For instance, the research in political science that follows Mayhew's (1974) classic analysis of rational political behavior of members of Congress largely informs legislators how they can best stay in office. One major purpose of this dissertation is to offer moral rather than practical guidance to political representatives. It should also serve as guidance to members of the public about what they should expect from their political representatives.

Like any other institutions, properly deliberative democratic institutions require the existence of *officials* who carry out the institutions purposes. In most institutions, and certainly in most democracies, public officials take oaths that give them a set of duties that they did not possess before taking office. It is uncontroversial that such *explicitly* acquired duties exist. However, the more interesting duties, and the ones I am concerned with, are not explicitly stated within oaths or contracts. They originate in the purposes of the officials' positions. These pur-

poses can be multifaceted, ill-defined, and often controversial; sometimes they may actually be indeterminate or incoherent. I will have to argue that they are implicit within the concept of an official's role in the democratic institution, and thus acquired upon taking office. This is decidedly harder than showing that duties listed explicitly in oaths are acquired by officials taking those oaths, but I will attempt to give a version of this duty that is weak enough to be necessitated by any adequate democratic institution. In the next chapter, I will introduce the "Duty of Political Responsiveness," and will attempt to suitably describe what exactly it requires of political officials, and especially political representatives within democracies. It will turn out to require that those political officials subject to it respect public opinion in some way appropriate to the position they possess and the institution in which they serve. In representative institutions the notion of 'respect' I will support requires political representatives to be "responsive" to public opinion. In the next chapter I will further explicate just what this means.

Chapter 3: Political Responsiveness

In this chapter my primary goal will be to describe and explain the significance of *political responsiveness* by political representatives. In the last chapter I argued that representative democracies must feature political officials who respect public opinion when making their policy decisions, and that this requirement is best based in the principle of a self-governance. It also constrains what can be meant by 'respect': it requires that political officials integrate public opinion into the policy decision process. This amounts to them being *responsive* to the public.

3.1 The duty of political responsiveness

At this point it will help to introduce a technical version of 'respect'. Label the specific notion of 'respect' that I will be concerned with as '*respect*$^{\text{O}}$'. This form of respect requires that public opinion be integrated into policy decisions. The rest of this dissertation will show how this integration can be made coherent.

First I will specify an institutional requirement for a system of government to be democratic: In order for representative democracies to instantiate the requirement of self-governance, and thus be appropriately democratic, the following duty must exist for some set of officials within the government:

Duty of Political Responsiveness *Representative officials* have a *prima facie* duty to *respect*^{O} public opinion when making policy decisions.

Within most political systems this will be a *moral* duty rather than a *legal* one. There are many reasons why it is best that certain duties possessed by political officials (as well as others) are not codified into the legal system; a moment's thought will reveal that the prospect of legal sanction for not being responsive to constituents may cause representatives to always act directly according to public opinion rather than just respecting it. Given what I have argued in the previous chapter, that would be undesirable.

One may certainly deny that any officials have the above duty, but as I argued in the previous chapter such a denial is not consistent with democratic decision making. I think that in most cases those who would deny the existence of this duty simply make a mistake about what it means. For instance, there are many public officials within modern democracies, such as judges and generals, who have traditionally made great effort to *ignore* public opinion. This is not a problem for my view, as I claim that there is only a duty of responsiveness for some set of officials; designate them "*representative* officials".[1] It is also consistent with this duty that some overriding considerations may give representatives reason to totally ignore public opinion; this could still constitute respect^{O} for public opinion.

Given what I have argued in the first chapter, I take the following to be a

[1]This is a broader category than "political representatives." That term likely only contains members of legislatures, but I conceive of representative officials as possibly also containing officials who are not explicitly granted representative and/or legislative authority, such as members of the Bureaucracy.

necessary requirement of any political system to be democratic:

Minimal Populism A political system can be considered democratic *only if* there is some class of public officials who have significant control over policy formation and who are generally obligated to abide by the *Duty of Political Responsiveness*.

This condition is distinct from a common definition of 'populism' as the expression of the popular will through voting behavior.[2] As I will show in section 3.3.1, I think that voting behaviors toward political candidates are poor measures of the "popular will" about public policies, at least insofar as "popular will" is equivalent to "public opinion".[3] I take "popular will" and "public opinion" to be collective *psychological* phenomena, and not merely behavioral ones. Therefore, my concept of minimal populism runs deeper than the standard view from voting. An early objection here may stem from this equivocation of 'will' and 'opinion'. I will explain why this is reasonable in the next chapter.

Given that *Minimal Populism* is only a *necessary* condition for a political system to be democratic, there are a host of other requirements that might also be necessary for any given political system to be considered democratic. Minimal Democrats place requirements on elections, and Deliberative Democrats likewise require that all policy be subject to deliberation and be contestable. Both of these

[2]I think this is typically how William Riker's influential definition of populism expressed in *Liberalism against Populism*, chap. 1 is interpreted, but it is more explicitly defined by Richardson (2002), pg. 58.

[3]Note that "popular will" is distinct from the "general will" that I discussed in the last chapter. "Popular will" is a more general term that is often used to express something like "policy preferences" in modern political science and philosophy. I will not seek to regiment it here as I have done with "general will."

constraints are consistent with *Minimal Populism*. If it is also assumed, as it often is, that only democratic political systems can be legitimate then this is also a necessary condition of a legitimate political system. I, however, will not defend *that* claim. In this chapter I will concentrate on a defense of *Minimal Populism* against a number of problems that arise when attempting to integrate public opinion into a decision procedure that also includes measures of expected well-being or interests, or some (other) measure of the common good. As I will show, the possibility and coherence of such a decision procedure is doubted by many; overcoming these doubts will require the development of a relatively sophisticated model of public opinion. Although particular details of this model may be shown to be incorrect without threatening my wider argument, the correctness of its basic features are vital for the plausibility of *Minimal Populism*.

Abiding by a duty to be responsive does not itself require that an official utilize some *particular* policy decision procedure. The duty as I have specified may admit a large set of possible decision methods. I will argue that an attractive, and perhaps best, model of this is a *weighing procedure* between public opinion and what the official takes to be in the public's best interest. However, this is not obvious on its face. Perhaps public opinion should act as a veto, or an initial screening device of policy alternatives. Additionally, deliberative democrats may argue for a role that is less straight-forwardly decision theoretic; for instance, they may require that public opinion sets the topics about which political officials deliberate. There is a further complexity that most policies are decided by some group of officials who represent different *constituencies* within society. Policy decisions have strategic elements as

well; if an official wants the government to adopt a policy most aligned with the opinion and interests of her constituents then it may sometimes be best for her to vote for policies that are *not* most preferred by (or in the best interests of) her constituents. Therefore, the decision procedure for a political representative may require elements of game theory as well as standard decision theory. I will largely avoid such issues, though will touch on them briefly in chapter 6.

The balancing procedure of policy decision is at least implicitly the orthodox view of political representation for liberal democracies. The most systematic description of this view comes from chapter 9 of Hannah Pitkin's classic work on political representation, *The Concept of Representation*. However, It must be emphasized that the balancing procedure is truly only *implicit* in Pitkin's work. She argues that most modern theories of political representation are faced with a conflict between requiring political representatives to follow the *mandate* given to them by their constituents and the *independence* those representatives have to do what they think is in the best interests of their constituents.[4] If representatives ignore their *mandates*, which require them to act in accord with the wishes of their constituents, then there is no real sense in which they are representing their constituents at all; rather they are ruling them, and an election is simply a tool to decide who should *rule*. On the other hand, if representatives ignore their *independence* and always act according to the wishes of their constituents then there is a sense in which they are not *representatives* at all, but simply are *agents* playing a similar representative

[4]There is slightly different language than that used in political science to describe the different behavioral "styles" of political representation, though they are largely equivalent to Pitkins distinctions. See Weßels (2007) for a discussion of how these different styles are categorized in political science. Achen (1978) provides a more formal presentation of representation styles.

role as a lawyer who files a lawsuit on behalf of someone or a police officer serving a warrant on behalf of the state. Agents use their best judgment only in cases when they are not explicitly instructed to do otherwise. However, their actions are always subject to direct control or overriding by clients. Pitkin argues that as the concept "political representative" is used it cannot be restricted to one or the other of these extremes. Rather, a political representative must balance between these two, using both the mandate given by her constituents as well as her own judgment about what would be best for her constituents.

Pitkin calls the balancing view of political representation *liberal representation*. This view takes its name from the "liberal" views of thinkers such as James Madison, especially in the *Federalist Papers*, rather than later "political liberals" such as John Rawls.[5] Madison described the two elements of political representation as a key to the new American legislative system, with regular elections making a *mandate* salient to representatives, while they retain broad legislative powers that permit them *independence*. Additionally, the two parts of congress were designed to operate at different points along the mandate-independence spectrum, with the Senate's indirect elections and longer terms permitting it more independence than the House of Representatives. Madison's primary concerns, however, were those of institutional design; he sought to give representatives incentives to abide by certain standards of political representation. My own concerns, and I think Pitkin's as well, are the proper actions of representatives *independent* of the incentives that may

[5]Of particular importance for this are *Federalist Papers* #51- #58, #62, and #63 .

exist within an institution.[6] For instance, someone who takes herself to be a representative of her constituency must surely be doing something against the norms of that office if she always acts in favor of those who can give her the larger campaign contributions. This is the case even within an institution with few safeguards against such actions. She is also likely acting *wrongly*. Of course, some legislative institutions are structured such that "constituents" have no expectation that their "representatives" will in fact be responsive to them. However, according to *Minimal Populism*, when no such institution exists to serve the role of responsiveness then that system of government should not properly be called a *democracy*.

An explanation of the mandate-independence spectrum does not make very clear how exactly political representatives should "respect" public opinion. I will argue that the balancing procedure should be modeled by a *weighing* (in a mathematical sense) of a representative's mandate against her independence.[7] I think that the other natural models of policy choice are unsatisfactory. Later in this chapter I will give a formal analysis of the weighing procedure that must be used to balance a representative's mandate against her independence. But it will be useful to first consider the obvious alternatives to the weighing conception of this procedure first; the reasons why such methods are not acceptable will shape how the weighing procedure should be formulated. The general theme that will run throughout my criticisms is that political decisions by political officials require relatively rich infor-

[6] Also see Brennan and Hamlin (2000) for an in-depth analysis of representation and institutional structure.

[7] The weighing conception of decision is well-known in decision theory, though often vaguely formulated. For a very recent discussion of the weighing conception, see John Broome's recent book, *Rationality Through Reasons*, pg. 52.

mation sources, whereas many standard decision procedures are rather information poor. The outputs of such information-poor procedures are therefore incapable of distinguishing between choice situations that intuitively should produce very different decisions. As I will argue, I think that only a weighing procedure between different *cardinal* measures of public opinion and common good is capable of adequately representing the sorts of decisions political officials are often faced with.

3.1.1 Other policy decision models

I will consider two kinds of alternatives to the weighing conception of the balancing procedure: "ordinal" aggregate models and deliberative models. First I will consider what I will call *ordinal* decision models. Some decision theorists object to the weighing (or balancing) of reasons in choice models.[8] One alternative to a weighing procedure is the use of only *ordinal* components in the decision procedure.[9] For instance, a political representative might include the fact that her constituents prefer some policy alternative x over another y and both of them over z in her policy decision, but not any measure of the magnitude of these preferences or the *differences* between those magnitudes. Such a limitation significantly constrains the role that public opinion might play within a policy procedure, and results in public opinion only modifying policy decisions in several rigid ways. Although there is a large set of possible ordinal decision rules that differ in their exact application and interpretation, for instance "*preemption*" and "*override*," they all share some

[8]For instance, see Horty (2012), pp. 2-5.

[9]Many systems of legal reasoning are compatible with purely ordinal forms of reasoning; for instance Joseph Raz's (1999) system of legal reasoning. However, most legal theorists do not consider in detail whether their system of reasoning is purely ordinal.

general problems.

Public opinion might *preempt* policy decisions by excluding rankings that some large proportion of individuals disagree with. For instance, if some large majority ranks a single policy alternative last in a set of policies under consideration (so prefer all alternatives to it), then that alternative may be excluded from consideration by the representative. Alternatively, public opinion may *override* the decision of a representative for similar reasons. Morality and law are rational systems that are often represented in this fashion, and my claims do not necessarily extend to those domains. But I do not think that the ordinal model is adequate for policy decisions. The primary reason for this is the way in which fine-grain differences in polices seem to matter quite a bit in policy decision.

Probably the most common examples of fine-grained policy choices come from cost-benefit analysis. Costs and benefits typically have clear cardinal significance, and the elimination of that significance seems to distort the overall analysis. Consider the set of policies x, y, z, and associate each alternative with a *cost* in number of lives lost over a year (or some other quantity) as a result of the policies: x:1000 lives lost, y:100 lives lost, z:10 lives lost. *Preferences* for these outcomes can be represented by the order $z \succ y \succ x$. But certainly the following statement about the sizes of the differences in preferences are also significant, where $u(\varphi)$ is the cardinal (numerical) expression of the preference for φ:

$$[u(y) - u(x)] > [u(z) - u(y)]$$

Note that if these numbers were simply subjective numerical expressions of preference — perhaps numbers individuals selected themselves to represent their cardinal preferences over the set of alternatives — then such a statement about the differences would not likely be significant, at least not without good theoretical reasons to think that they can be interpreted in this fashion. But lives lost have a natural cardinal significance, and it is unreasonable to think that an ordinal model of balancing could adequately represent an interaction between public opinion and such a cost-benefit analysis.

Using the above example, it can easily be seen what makes an ordinal balancing procedure problematic. Assume that the public preference over the policies can be represented by the ranking $x \succ y \succ z$. Furthermore, to make matters simpler, stipulate that this is a *near unanimous* ranking. How might representatives use this information? As I already stated, they might exclude z from consideration, because it is ranked lowest by the public. Or because there is near unanimous agreement among the public that x is the best, they may simply choose x. I think that such a decision is impossible or will implicitly involve cardinal information of public opinion. The first question I think representatives would ask themselves is about the *strength* of the public favoring x over y; if there are stirrings in the public of a very strong preference of x over y (and I will address below how this might actually be measured), then it might be reasonable to choose x at the cost of 900 lives. But this would seem to be a horribly misguided choice if x were barely (though nearly unanimously) favored over y. I think a careful reflection on different domains in which policies must be chosen will reveal a similar structure.

The ordinal methods of balancing that I have described also implicitly introduce cardinality in the conditions of the application of public opinion in the decision (or the application of whatever criteria are balanced against *public opinion*). For instance, public opinion about some alternative pair might only override the judgments of officials when more than 70% of the public agrees on the ranking of those alternatives. But this in itself constitutes an estimate of some sort of difference of magnitudes of *social* preferences, albeit only two classes of differences (those that are greater than or equal to the threshold and those that are lower). The advantage of this level of cardinality is that it is clear where the magnitudes come from; they are just the different numbers of people supporting various rankings of the alternatives. But I think this returns to the first problem I suggested with the ordinal balancing method; it does not provide the sort of information that seems necessary for policy decisions. For the reasons I give above, just two classes of preference magnitude differences are not enough to balance against the sort of cardinal information common in cost-benefit analyses.

One might attempt to infer more classes of preference differences from information about the number of people who support a given preference ranking. This just requires the inference of social preference magnitude differences from the number of individuals with various social orders, which is a different method for obtaining the cardinal information about preferences than what I will present here, but nonetheless it abandons the ordinal framework. However, such a method has implications for the practical implementation of the sort of public opinion measurement method I suggest here, and so I address it in section 6.2.1.

The second category of alternative decision procedures comes from deliberative democratic theory, which I introduced in section 2.1.3. One interpretation of Habermas' discursive democratic theory, for instance, requires that informal deliberation in the public "flows" up to the formal deliberation process amongst political representatives. I generally accept the basic structure of this view, but many deliberative democrats may be committed to thinking that this information flow can only be in the form of deliberatively relevant information: primarily *arguments*, *reasons*, and *topics* of importance. This is the aspect of the procedure that will be problematic if it is hoped to allow for self-governance, which I have argued is necessary for democracy. A second related type of deliberative process comes from Cohen's concept of "radical democracy", where general items of concern are decided on by political representatives, but details are decided by deliberations amongst members of the public.[10] The "radical" feature of this view is the fact that most particular decisions are made within deliberative sessions among the *public*. Political officials are primarily tasked with a certain amount of direction of public deliberation, which might also be construed as *agenda setting* (to use a term from social choice theory).[11] This notion of deliberative democracy is clearly *democratic*. Additionally, it seems to satisfy *Minimal Populism* as I defined it above, because people are in a sense their own representatives. Therefore, it will be instructive to see how a system of government that satisfies minimal populism might be undesirable for other reasons; most importantly, for being too populist.

[10]See Cohen and Fung (2004).

[11]In social choice theory, agenda setting is typically taken to be a form of manipulation. However, one might also think of it as important to any public deliberation so that the most important issues are discussed. This is its use in "radical democracy."

Both of Habermas' and Cohen's decision procedures integrate public opinion into the decision. Habermas' procedure permits public opinion to form the backdrop against which deliberation by political officials occurs, whereas Cohen's view makes deliberation by the public the primary source of eventual decisions. As I argued in the last chapter, I think that Habermas' view gives far too little respect to public opinion as a source of considered policy judgments of the public, whereas Cohen's gives far too much respect to those judgments. Referencing the goal of the general will that I introduced in the last chapter, Habermas' procedure will fail to get at the general will, because the procedure fails to find its source in the sovereign will of the people. Habermas' deliberative "flow" is only concerned with the arguments and reasons given by the people and not with their considered judgments (their opinions). Therefore, it fails to instantiate the general will.

Argument forms and reasons for accepting the truth of their premises need not be attached to individuals, at least not when they are publicly acceptable (and thus admissible within deliberation), and so wills of the people themselves never enter into the deliberative procedure itself. This is much like Edmund Burke's conception of political representation, in which representatives are thought to present the interests of their constituents as *abstract* and therefore fully detached from the individuals who have them.[12] But Burke's view is most decidedly non-democratic, at least to modern readers. I think Habermas' view suffers from a similar problem, but perhaps an even worse case of it than Burke's view. In rational deliberation, arguments

[12]See Pitkin's (1967) discussion of Burke's views on representation in chapter 8 of *Concept of Representation*. Burke does not present a single coherent picture of political representation in any of his writings and therefore it is rather difficult to piece together without the help of secondary sources.

and reasons that originate from the public are not given any special status; all rational considerations must be treated equally for the process to be considered rational. When this is the case, and political officials rather than the people are in the position to judge the relative merits of various arguments and reasons, then small but articulate and intelligent private interest groups will often deserve more support than much larger groups of citizens who possess lesser rational capabilities. From a purely deliberative standpoint this is a desirable outcome, because the better argument expressed by a tiny majority is able to win the day. However, when such arguments are the only means by which the people are able to affect the decision process, the system can rightly be called democratic no more than Burke's view of political representation.

Cohen's view of radical democracy obviously meets the requirements of *Minimal Populism*, but is inadequate for another instructive reason. In the radical democratic framework most important political decisions, especially those of local importance, are made by deliberative groups of average citizens. Even assuming that none of the standard practical problems with such massively deliberative systems arise in this case, Cohen's democratic model fails to take into account the role that political officials have in using their expertise and position within a professional deliberative body to point the direction of policy toward the public interest. This directedness is important to avoid policy decisions that merely aim at some aggregate of preferences of the members of society, something akin to the *will of all*, rather than the *general will*. Additionally, local deliberation will never be able to resolve issues of regional or national concern. This will, it seems, again require

deliberation among political representatives, given that in modern societies large national programs are particularly important for the lives of citizens.

I have considered both decision theoretic alternatives to the weighing procedure (preemption and overrule) and deliberative alternatives. However, I have not presented any details of what a weighing procedure involving public opinion might look like. In fact, there are likely many ways by which a weighing procedure could be modeled. In the next section, I will provide what I think is a reasonable formal representation of the procedure by which political representatives might weigh public opinion against their estimates of what is in their constituents' best interests.

3.2 A model of policy choice

My purpose so far has been to show the necessity of a weighing conception of policy decision by political representatives. For representatives to respect$^{\mathrm{O}}$ public opinion, they must use a weighing procedure that utilizes cardinal information about both public opinion and their constituents' well-being (as estimated by the representatives). Next I will present a formal specification of such a weighing procedure consistent with the *Duty of Political Responsiveness*.[13] Specifically this is a model in which both public opinion (representing the mandate) and the expected welfare produced by the policy (representing independence) are weighed against one another. I will use expected welfare measures as a stand-in for a measurement of

[13]The duty is general enough such that there are many specific models likely comply with it. Therefore, there may be better models than the one I provide here. However, I think the one I present here is intuitive and provides a suitable test-case for the duty and exhibits a number of important problems that I think will be common to most possible decision procedures.

what is in the common good. I do this for convenience, and because welfare is a concept that is (relatively) easily represented with expected utility measures. Note that the model I present here only covers a subset (public opinion and expected welfare) of the total set of possible decision criteria relevant to policy choice. Morality, for instance, may provide either weight for some policies or act to override the result of such a weighing procedure.[14] Therefore, the model I present here should only be thought of as one piece of a potentially much larger decision procedure.

The following are the basic elements of a procedure that weighs public opinion against an official's beliefs about what is in the interests of her constituency.[15]

$\mathcal{I}$ *Constituency* of the official, which contains a set of individuals $I_1, I_2, I_3... \in \mathcal{I}$.

F_i Set of *policies* available to the official's choice, where alternatives $x, y, z... \in F_i$

O_i Set of *outcomes* that could result from the choice of a policy, $o_1, o_2, o_3... \in O_i$

$\mathcal{P}(F_i)$ Set of subjective probability distributions of the official over the set O_i, for each member of the set of policies F_i by the official. For example, $P_x(O_i) \in \mathcal{P}(F_i)$ is the probability distribution that results from the adoption of the policy x over the set of outcomes O_i

$\succ_{F_i}^{\mathcal{I}}$ The collective preference function of individuals in $\mathcal{I}$ over the alternatives within F_i.[16]

[14]This is especially likely if one uses a non-utilitarian moral theory, because there will need to be some way to account for the rights of individuals that are not themselves based upon utility or preference. This is a sort of "laundering" of the *outputs* of the social decision (Goodin, 1995).

[15]One could represent this with any standard decision theoretic model, for instance that of Anscombe and Aumann (1963), though that would require some work.

[16]Note that I will use 'alternative' as a general technical term that can be any item within a preference function.

$\succ_{\mathcal{O}}^{\mathcal{I}}$ The official's estimate of the preference function for individuals in $\mathcal{I}$ over the outcomes within O_i based upon well-being.

The set of alternatives that is immediately under consideration by an official (F_i) is only a subset of all possible alternatives. This will be adequate for an explanation of the basic model, but in chapter 5 it will be necessary to consider the entire set of alternatives under consideration within the same choice *context*. I will explain choice contexts in more detail later, but a choice context can be thought of as a set of policies about a similar topic or attempting to solve a similar problem. For instance, a set of possible alternative health care systems would likely be contained in a single choice context, and a set of policies that advocated various foreign policies would likely be contained in another. Later, I will require that all alternatives within a single context could theoretically be placed along a single cardinal utility scale, something that may be impossible *across* choice contexts. Therefore, define a set of terms that allows for representation of choice contexts:

$\mathcal{F}$ The set of all possible alternatives that could be under consideration (over all possible choice contexts).

D_i A choice context, such that there is some set $D_i \subseteq \mathcal{F}$

$\mathcal{D}$ Set of all choice contexts.

$\mathcal{O}$ Set of all possible outcome sets O_i.

In this model every official has some *constituency* ($\mathcal{I}$) that serves as the information base for all questions of public opinion and interest. The identity of the

constituency may depend both upon explicit norms within the political system as well as a more comprehensive theory of political representation. Within most democracies, explicit norms of constituent identity typically track the set of individuals eligible to vote for the representative in elections; so for instance, a U.S. Senator has an explicit constituency that contains all adult citizens of her U.S. state, whereas a member of the House of Representatives only has an explicit constituency containing adult citizens in her house district.[17] However, in the U.S. there seems to exist a non-explicit norm that makes the public opinion and interests of *all* individuals within a representative's own state more relevant to her decisions than those outside her state. Given that my model here is not restricted to political representatives in legislatures, it may also include constituents of various government agencies and officials within those agencies. Although it is true that most agencies in principle serve all citizens or residents served by a particular government, it would be unlikely that many agencies would make such broad inquiries into the opinions and interests of all those individuals. Rather, agencies are typically called to pay special attention to those individuals who have significant stake in the policy decisions the officials make.

Other more foundational reasons might lead officials to take certain views about contents of $\mathcal{I}$. For instance, someone who holds the Burkian view of political representation (which I mentioned earlier) would hold that political representatives

[17]Here I ignore the further complexity presented by children, incapacitated persons, or those who are otherwise ineligible to vote. In the American context, the most interesting are the cases of non-citizen residents and felons who have lost the right to vote. Even though such individuals cannot vote, their interests and opinions seem to command *some* weight in the decisions of public officials.

should not privilege the interests of the members of their districts, but rather should always consider the interests of all members of the state. Although Edmund Burke thought that an important role of a representative was to represent the interests of the members of her district within the legislature, this did not mean that the representative should pursue those interests at the cost of the interests of those outside her district.[18] Therefore, a Burkian concept of constituency is far broader than the concept typically employed in American politics. Phenomena such as those above hint at a complex weighting procedure that officials should use when determining the relative influence of different individuals on their decisions. However, such a procedure could be added to the model I provide here, and does not seriously hinder the argument I will make.

Every *constituency* ($\mathcal{I}$) is represented as having a social preference function over various sets of alternatives. Any constituency has such functions over many different sets of alternatives within different choice contexts in the set of all possible alternatives $\mathcal{D}$; therefore, it need not be assumed that all possible alternatives are comparable. *Outcomes* (in $\mathcal{O}$) can be defined in a multitude of ways, though it is likely simplest to think of them as descriptions of possible worlds. And although these outcomes must be described in a fashion detailed enough to adequately determine their effects on the constituency, and thus the resultant social welfare function over outcomes, they need not be *completely described* possible worlds. Many differences in possible worlds simply make no difference to welfare. Additionally, actual

[18]It should be noted that Burke would have denied that public *opinion* should ever be considered by political representatives in making decisions.

utility estimates (such as those from cost-benefit analysis) are never capable of differentiating meaningfully between the relative goodness of possible worlds that are too similar, nor the relative probabilities of possible worlds given each policy alternative. Any actual analysis will be a crude estimation of the expected utility of some policy.

According to the concept of representative *independence*, political officials are in part tasked with determining the most likely set of probability distributions ($\mathcal{P}$) given the set policy alternatives under consideration (F_i). *This* aspect of the model is not likely to be controversial, though what specific use officials should make of such a set of probability distributions is certain to be. *Figure* 3.1 shows a standard choice situation for some official. Each policy ($x, y, z \in F_1$) is associated with some set of outcomes (O_1) and a probability distribution over that set ($\mathcal{P}(F_1)$) . In the figure, this distribution has been reduced to probability values over the three possible outcomes ($o_1, o_2, o_3 \in O_1$), but in any more realistic case the set of possible outcomes will be far larger. This small set of possible outcomes means that ($\mathcal{P}(F_1)$) can be described in terms of three probability values over the set of possible outcomes for each policy alternative ($\{p_i, q_i, 1 - (p_i + q_i)\}$).

Given what I have already assumed, officials have at their disposal preference functions by the public for both policies *themselves* (public opinion) as well as the outcomes of those policies (expected welfare).[19] There are a number of ways in

[19]Harsanyi (1982) theorizes about a similar distinction between *utility* derived from "moral preferences" and "personal preferences." He suggests that an individual may have a "moral" preference for some policy that will nonetheless produce worse outcomes for that particular individual (an outcome ranked lower by that individual's "personal" preferences). For instance, the wealthy person may prefer higher taxes on the wealthy in her set of *moral* preferences even though she

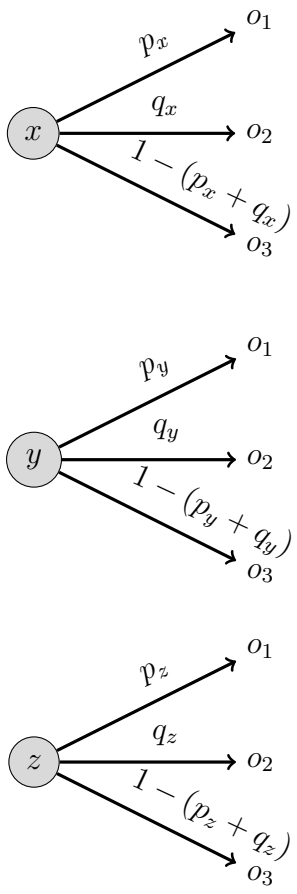

Figure 3.1: Expected welfare

which officials could use this information. Consider the method that relies entirely upon the direct preferences of the public for the policies, thus ignoring the contents of $\succ_{\mathcal{O}}^{\mathcal{I}}$ and $\mathcal{P}$ entirely. This would be similar to Cohen's "radical democratic" view (Cohen and Fung, 2004), whereby officials make the components of their welfare calculations known to the public during some public deliberation on the topic, but policy decisions are based upon public preferences alone. In the way I have formal-

recognizes that her and her family (which she may care most about) will be worse off under such a tax system, thus her *personal* preferences rank higher taxes lower than the status quo. Harsanyi thinks that the *moral* preferences of individuals will tend to assign equal weights to the well-being of all individuals within the society, and thus will also tend to be *identical*. Although this last claim does not seem plausible, Harsanyi's general point is useful. Individuals possess different criteria for evaluating policies, which produce distinct preference sets. And individuals are capable of using these different preference sets within different choice contexts.

ized the decision process here, such deliberation would just affect the contents of $\succ_{F_i}^{\mathcal{I}}$ and not the decision process itself. Using such a method, the social decision is some social choice function over F_i, utilizing only the information contained within $\succ_{F_i}^{\mathcal{I}}$; this only abides by the *Mandate* given to officials by their constituents.

Alternatively, officials could ignore $\succ_{F_i}^{\mathcal{I}}$ and instead only use $\succ_{\mathcal{O}}^{\mathcal{I}}$ and $\mathcal{P}$. This would maximize the *Independence* of officials, if it is assumed that the sole end for which they should use their independence is to further the collective welfare (or common good). This could also be reduced to a Harsanyi- style utilitarianism if the officials used some averaging (or summation) function to determine the best outcome. However, officials could also use a more complex function, such as *maximin*.[20] The essential feature of their independence is not a specific choice function used to make the decision, but rather the use of the expected welfare measure rather than *policy* preferences to make policy decisions. Also note that these outcome preferences could take many forms, including measures of objective well-being, preferences over the outcomes, counts of Sen-style capabilities or Rawlsian primary goods, or the satisfaction of idealized preferences.[21]

It is not my intention to provide a complete explanation of how the expected welfare of the constituents should be measured. In fact, not much should even be read into my choice of 'welfare' to designate the subject of officials' *Independence*. I will largely ignore a number of important issues that arise when welfare or a similar concept is used as a criterion of choice, but I think these can largely be

[20]The maximin function maximizes the utility of the individual with the least utility.

[21]The "capabilities approach" and "primary goods" serve central parts of Amartya Sen's and John Rawls' views (respectively) of how well-being can be represented and compared across individuals. See Sen (1992), chapter 3 and Rawls (1971) chapter 2, section 15.

compartmentalized. For instance, it is likely that officials should include future welfare of members of the society in their calculations of welfare, but the future welfare should be discounted against current welfare. How the future should be discounted is a troubling question that any consequentialist view of policy choice will likely encounter, but I only require that the measurement of welfare take certain structural properties that permit its balancing with the policy opinions. And such structural properties will be present regardless of how different sources of welfare information are integrated to form the measure of total (or all-things-considered) welfare.

Complete *Independence* for the official would require the public decision to be fully determined by what that official believes to be in the best interests of the constituents, assuming that it was the prescribed method to determine the interests of the constituents. This dissertation is primarily a study of the role of *public opinion* in decision making of public officials, and so I treat $\succ_{\mathcal{O}}^{\mathcal{I}}$ primarily as a placeholder. However, several of its features will be important for my analysis. It must first be shown that some social preference function over F_i can be calculable from $\succ_{\mathcal{O}}^{\mathcal{I}}$ and $\mathcal{P}$ in order for the expected welfare of the public to be weighed against public opinion. This is possible so long as $\succ_{\mathcal{O}}^{\mathcal{I}}$ is a cardinal utility function. Every alternative in F_i is associated with a position in $\succ_{F_i}^{\mathcal{I}}$ as well as a set of probability assignments over the set $\mathcal{O}$. These probability assignments and the preference function $\succ_{\mathcal{O}}^{\mathcal{I}}$ allow for a set of utility assignments over F_i to be calculated that meets the conditions of *expected utility theory*, as long as it is assumed that it is meaningful to use the probabilities P in such calculations in the first place. Label this new function:

$\succ^{\mathcal{I}}_{\mathcal{O}\times\mathcal{P}}$. In other words, officials are able to find a *cardinal*-value utility function over the set of policy alternatives under consideration (F_i) that represents the official's best estimates of the expected welfare that would result from the adoption of those alternatives. Remember, however, that this need not be a *simple* expected utility function; it could be something more complex. However, for any standard utility function that utilizes the probabilities found in P, it must be assumed that the resulting function meets the requirements of expected utility theory, giving it cardinal measureability. Therefore, the *weighing* model can be described very simply. Where there are two positive weights w_{F_i} and $w_{\mathcal{O}}$ on $\succ^{\mathcal{I}}_{F_i}$ and $\succ^{\mathcal{I}}_{\mathcal{O}\times\mathcal{P}}$ respectively, the social decision function $f(\varphi)$ that balances the mandate and independence of the political representative to find the utility assignment on (the alternative) φ is found by:[22]

$$f(\varphi) = w_{F_i} \times \succ^{\mathcal{I}}_{F_i}(\varphi) + w_{\mathcal{O}} \times \succ^{\mathcal{I}}_{\mathcal{O}\times\mathcal{P}}(\varphi)$$

As a formal matter of *calculation* this weighing procedure does not necessarily require that either public opinion or expected welfare are along cardinal scales. Weighing could be allowed if one or both of the scales are along *ordinal* scales through something like a majoritarian social choice procedure. Assume that both functions have ordinal measureability and the weights are positive *integers*, weights can then be given the role of the number of "voters" possessing the preference functions. So if the interests of individuals are given double the weight as public

[22] $\succ^{\mathcal{I}}_{F_i}$ and $\succ^{\mathcal{I}}_{\mathcal{O}\times\mathcal{P}}$ must both be associated with utility functions. However, if either of these utility functions only have ordinal measureability then the result of $f(\varphi)$ will be arbitrarily determined by the particular way in which utility assignments are made. This is a formal way to understand the requirement of cardinal measureability on both measures of public opinion and expected welfare.

opinion in the weighing procedure, and:

$$\succ^{\mathcal{I}}_{\{x,y,z\}} = x \succ y \succ z$$

$$\succ^{\mathcal{I}}_{\mathcal{O}\times\mathcal{P}} = z \succ y \succ x$$

The resultant ordering must be $z \succ y \succ x$. In fact, the resultant ordering of the weighing procedure with this set of weights must always be equal to the ordering of the constituents' expected welfare. And this example isn't unique; when there are only two criteria, *any* weighing will result in the direct adoption of whichever component has greater weight. This results in a dictatorship of the greater weighted criterion. As I argued earlier in this section, the weighing procedure must be sensitive to magnitudes of utility differences if the procedure is to allow for large differences in criterion to overcome smaller differences in the other criterion, even when the criterion with the smaller differences is weighed more than the one with large differences. One complication of the cardinal weighing procedure is that it seems to require that the different cardinal scales (public opinion, welfare, etc.) are comparable. I deal with this in detail in section 6.1 when I present the full weighing procedure.

The demandingness of the weighing conception may be seen by some as a very good reason to reject it. However, I think that such a conclusion should be resisted. I showed in section 3.1.1 that both ordinal and deliberative uses of public opinion are unsatisfactory. Furthermore, the weighing conception has a great deal of intuitive appeal, and is very influential in our everyday notions of how political

officials should make decisions. It is certainly easy to imagine political officials weighing magnitudes of public opinion against what will produce the greatest welfare of their constituents during their decision process. It would seem odd that this reasoning process is illusory or otherwise incapable of being modeled. And, as I have already claimed, the common measures of welfare are often cardinal. A cost difference between two policies of $4,000 is clearly significantly larger than $100, and a difference in the loss of lives of 400 people is significantly larger than a loss of 10. Exactly how these differences should be factored into the composite welfare or interests measure contained in $\succ_{\mathcal{O}}^{\mathcal{I}}$ is not obvious, but I see this as exactly the problem that modern cost-benefit analysis takes itself to address. Therefore, there is reason to pursue the possibility that public opinion can be represented by using *cardinal* measures , and that it is capable of being modeled as such. However, any use of public opinion faces a number of serious problems, both from empirical research in political science and conceptual research in social choice theory. In the next section, I will evaluate a number of ways in which public opinion is measured in large representative democracies. I conclude that most current methods produce only *ordinal* representations of public opinion, and are thus not adequate for the weighing conception of policy choice. Furthermore, public opinion appears to suffer from serious problems of preference *instability*. This has made many question whether it can be modeled with standard preference and attitudes representations that assume preference stability. In the next chapter, I will begin a presentation of a public opinion model that does meet the requirements of the weighing conception.

3.3 The problems with measuring public opinion

If the political *Mandate* is to be taken seriously, then there must be some way that political representatives come to know the *contents* of that *Mandate*. In the next sections I will discuss a number of methods commonly employed by political representatives (and others) to learn about the public opinion. I will use social-choice theoretic analyses to show that these methods cannot possibly provide political representatives with the information necessary to include public opinion within a weighing procedure. However, the potential problem with public opinion runs deeper than the methods currently used to measure it. A number of findings in political science indicate that many members of the public are relatively ignorant about many facts relevant to public policy debates. Additionally (and perhaps because of their ignorance), most members of the public seem to exhibit "unstable" policy opinions, which seem to change randomly through time. Consequently, it is of great controversy in political science just how well-formed and well-*in*formed public opinion is; I will discuss this debate in section 3.4.

3.3.1 Voting

The 2008 U.S. presidential election had the highest national voter turnout rates of any national U.S. election since the 1960s. For democratic theorists who take voting to be the best measure of national public opinion, this seems to have been a high water mark of public opinion measurement. Nonetheless, in the general presidential election of 2008 only about 62% of eligible voters cast a ballot for

president.[23] And even on that day, voter turnout in many states failed to surpass half of eligible voters. If election results are taken to instantiate in some way the opinions of the people — *all* of the people — then actual voting behaviors seem to put into serious question the foundations of the voting–centered concept of the political *Mandate*.[24]

The voting apologists may attempt to exclude the preferences of non-voters. One classic interpretation of voting behavior represents a vote as an indication that a threshold of utility that voter associates with a given election result has been met.[25] This overcomes worries that voting is never rational, given the small chance that any given individual's vote is decisive. When an individual votes, this counts as a measurement of their positive utility achieved from their act of voting, based upon the likelihood that their vote is decisive, minus the negative utility of the act of voting itself. If the negative utility associated with voting is known for each individual, then voting does have some potential to provide some genuine information about the utility each individual attributes to the attainment of the public decision at stake. There have been criticisms of this general economic analysis of voting behaviors,

[23]This data comes from the *United States Election Project* (2012).

[24]There is a very old idea in American political thought, expressed by Jefferson and others, that there is no expressed right to have one's policy opinion included in policy deliberations, but rather only have a right to an *opportunity* for such inclusions. Such a view would permit the exclusion of many individuals from the measure of public opinion if they have not actively taken part in the political process. This in part stems from the more literal active interpretation of the public "will," according to which the will of the people is some collective *action*. This contrasts with the psychological interpretation of the will that I have provided. A similar view, though one not reliant upon a concept of the public will, is expressed by Pettit's republicanism (1997), in which members of the public need only be given opportunities to contest the actions of government in order for it to be democratically legitimate. It was these views that I attempted to cast doubt on in chapter 2.

[25]See Downs (1957) for a classic utility-based view of voting decisions.

likely the most important coming from "expressive voting" theory.[26] This view takes voting behaviors as expressions of underlying individual policy preferences and not underlying utility analysis by the individual. However, both the expected utility and expressive voting views are susceptible to similar formal criticisms.

It can rather easily be seen that a large component of what determines voting turnout behavior is not related to the utilities individuals assign to the various social outcomes that would result from alternative candidates. For instance, in the U.S. there is a very systematic oscillation in voting turnout in national elections, with peaks occurring during years with a presidential election, and troughs occurring during non-presidential years.[27] One may interpret this oscillation as an indication that people are generally more concerned with the outcome of presidential elections than with non-presidential elections; this does have some intuitive appeal given the importance people in the U.S. assign to the office of President. However, such an explanation would present a number of additional problems for interpreting voting behavior in *congressional* elections. Consider a member of the house of representatives attempting to interpret the *Mandate* produced by the most recent election. Assume that there is a midterm election E_m, with a given policy platform $\succ_m$ that ranks some set of policy alternatives F_1. The election two years prior to this was a presidential election year (E_p) and had a different platform of $\succ_p$ over the same set of alternatives F_1. Of course, in the simplest case, $\succ_m = \succ_p$, but this need not always be the case. It is unclear whether $\succ_m$ can ever provide a mandate if it is

[26] See Brennan and Lomasky (1997).

[27] For instance, the non-presidential years of 2002 and 2006 had national turnouts of 29.5% and 48.9% respectively, whereas the presidential years 2000 and 2004 had turnouts of 54.2% and 60.1% respectively.

different from $\succ_p$. Therefore, assume that $\succ_m$ and $\succ_p$ significantly differ (however that is to be measured), and that E_m has half of the voter turnout of E_p, with the representative winning easily in both cases. It becomes quite unclear at this point what exactly is the representative's *Mandate*. Perhaps the most reasonable thing to say, and what is often said in politics, is that *clear Mandates* (those that should certainly be acted upon) only arise in elections in which a candidate wins a convincing victory with high voter turnout. The election results of E_m might simply signal a great deal of indifference within the electorate, thus permitting the representative to act according to her *Independence*. In this way, E_m might have the effect of wiping away the previous *Mandate* from E_p though not bringing with it a new *Mandate* of significant strength. But this in fact just undermines the ability of an election to measure the *Mandate*.

The waters get even muddier when considering the many reasons people vote for various candidates. In the above analysis I assumed that people only vote for candidates based upon policy platforms, but people may also vote for candidates who they think will *govern* well. Given that people typically have better formed opinions for political leaders than possible public policies (Converse and Markus, 1979), it may also be *reasonable* that they do this. And this is part of Schumpeter's famous criticisms of populism: given how unclear the connection is between the election of candidates and the adoption of public policies, it seems reasonable that the public vote for candidates they trust to govern well rather than just those who have policy platforms of which they approve.[28] Additionally, governance in

[28]See Schumpeter's *Capitalism, Socialism, and Democracy*, especially pp. 256-282.

the American context is often thought to be a *competitive* exercise, in which players (the political representatives) compete for influence within the government. For this reason, self-interested voters — those who want their preferred policies instituted — should not necessarily choose the candidate who matches her policy preferences most closely. This is made obvious by a simple example. Consider a set of policy alternatives $\{x, y, z, a, b\}$ of a voter with a strict *priority* ordering of those alternatives $x \succ y \succ z \succ a \succ b$. This means that she would prefer that x obtains more than she would prefer that y obtains, and so on.[29] Consider three candidates for public office, each of which have an agenda and some probability of enacting that agenda based upon the political *skill* of that candidate (where $\neg x$ stands for "x not obtaining"):[30]

C_1 60% chance of enacting $\{x, \neg y, z, a, \neg b\}$ if elected

C_2 30% chance of enacting $\{x, y, z, a, b\}$ if elected

C_3 70% chance of enacting $\{\neg x, \neg y, z, \neg a, \neg b\}$ if elected

If it is expected that a voter will choose a candidate that is most likely to enact her policy preferences, C_3 should be eliminated right away, because that candidate has a high likelihood of enacting policies that the voter disprefers.[31] It is unclear what the voter would decide in the comparison between C_1 and C_2. It is true that

[29] We also assume that, for example, if x obtains then $\neg x$ does not obtain.

[30] We might might also include the trustworthiness and other features of each candidate that determine the likelihood of each enacting their stated platforms.

[31] Note that there too little information in this example to say what it is *rational* for the voter to do this. That would require having the voter's cardinal utility assignments over the alternatives. However, my example here shows that there are factors that go into the voters decision other than each candidate's platform.

C_2 will attempt to enact identical priorities to those of the voter, but C_2 is less likely to be able to enact the voter's highest priority policy (x). In such a case, the voter must decide whether she is willing to trade any possibility of enacting y or b for a higher probability of enacting x, z, or w.

If a voter's preferred platform is not the only consideration that voters should take into account when choosing a candidate, then it must be assumed that members of the public do not necessarily act rationally if they vote for candidates on the basis of platform alone. It is likely that voters act irrationally about a great many matters, but not likely that this irrationality makes them *more* likely to express their policy preferences in terms of their voting behavior. Because candidates must be elected in order to be involved in the enactment of any policies, voters might also only vote for candidates whom they believe have a chance of winning, so that they are not "throwing their votes away."[32] In any case, this need not be resolved here. I have shown additional considerations voters may use to make a voting decision besides a candidate's platform, but this is only one among potentially many others. Every additional consideration voters might use to make voting decisions beyond the policy platform of the candidates further alienates the voting choice from the policy preferences of voters. If the plausible assumption is made that these factors have significant effects on the voting behaviors of a non-trivial percentage of voters, then it would seem that elections are a poor source of information about public opinion.

[32]In social choice theory this is called *strategic voting*, and there is no voting method that is not susceptible to it (Gibbard, 1973; Satterthwaite, 1975).

3.3.2 Direct communication

Perhaps the most direct method by which constituents can express their policy preferences and the reasons for those preferences is through direct communication with their representatives. Consequently, the obligation of a representative to listen and respond to her constituents in a variety of ways is ingrained into the practice of American democracy, for both its function in making public opinion known to representatives as well as aiding in public deliberation. And in fact, such direct communication is one of the main activities taking place within American Congressional offices.[33] When a constituent communicates directly with a representative or the representative's staff — be it in person, over a telephone, writing, or through protest activities — it is intuitive that some important feature of a constituent's attitudes are expressed to the representative. Unlike voting, direct communications from constituents do seem to give some clear information about the policy opinions of the constituents themselves; it is this capacity of communication upon which I will concentrate. A deliberative democrat will be more concerned with the *arguments and reasons* behind her constituents' policy opinions, and these are frequently the subject of direct communication. I don't doubt the usefulness of such exchanges for deliberative politics, but I will present some doubts about its usefulness for understanding public opinion.

In order to investigate the ability of such direct communications to convey preferences useful to a representative, I will attempt to model them in the social

[33]See Mayhew (1974) for a discussion of the major portion of congressional activities that involve direct constituent communication and services.

choice theoretic terms. Although such a framework is more natural in the case of elections, I think that a similar analysis will provide clarity about the structure of the information provided by direct communications. Consider a set of four communications from three members of a constituency, $\{I_1, I_2, I_3\}$ within a constituency of 50 total members:

1. I_1 writes a letter stating that, "I wholeheartedly support the new plan to give every American free health care, and think that a 1% increase in taxes is small price to pay."

2. I_1 calls the office of her representative, reiterating her support for free health care that she recently expressed in a letter.

3. I_2 stands outside the office of the representative every day for a week, holding up two signs, one saying, "Keep your hands off of my health care" and another saying "Taxed to death."

4. I_3 signs a petition against any increases in taxes from new health care legislation.

A representative must first determine whether to treat communicating constituents differently from non-communicating ones. Given that a very small portion of constituents communicate with representatives, it is unlikely that a representative should use communication as the *primary* means of gaining information about public opinion. However, one might argue that communication provides some useful information, especially when a representative receives a great deal of communication

about a particular topic. In such a case, communication might provide a sample from the population of all constituents. Certainly this is not a *random* sample, but it might be a sample selected based upon the *strength* of opinion. Under such a view, an individual only communicates with a representative once the extremity of the *utility* she assigns to the alternative under consideration reaches a certain threshold. Therefore, the amount of communication that a representative receives might provide information about the preferences of the entire population of constituents. This is similar, then, to the threshold view of voting that I discussed above.

If the threshold view of political communication I describe above is assumed, it may be possible to state something about the cardinal values of the utility each individual assigns to the policy alternatives under consideration (F_i). For example, define two policy alternatives: universal health care and higher taxes (h) and no change in either healthcare or taxes ($\neg h$). Then, for each individual $I_i \in \mathcal{I}_C \subset \mathcal{I}$, and each individual I_n in the set of non-communicating constituents $\mathcal{I}_N \subset \mathcal{I}$ (where the union of $\mathcal{I}_C$ and $\mathcal{I}_N$ exhaust the space of individuals in $\mathcal{I}$):

$$\forall(i \in \mathcal{I}_C, n \in \mathcal{I}_N)(|u_i(h) - u_i(\neg h)| \geq |(u_n(h) - u_n(\neg h)|)$$

This just says that the amount of utility *difference* each member of the communicating group assigns to two alternatives is at least as large as that assigned by those members in the non-communicating group. This may even be a strict inequality, though that would be a stronger claim and wouldn't change much about my conclusion. Importantly, this statement does not specify *how much* greater the

utility differences are for those who choose to communicate and those who do not. It may be the case that those in $\mathcal{I}_N$ are largely indifferent between the alternatives, such that the utility differences from the three individuals within $\mathcal{I}_C$ could overwhelm the differences from the 47 within $\mathcal{I}_N$ if a simple averaging procedure were used to calculate the social utility of adopting h. However, these quantities are unknown, and it seems likely that it would be impossible to obtain such information by simply communicating with constituents; this is exactly the problem of subjectivity in cardinal utility measures that concerned Arrow (1963) in his limitation of standard social choice theory to ordinal measures. I will discuss that problem in more depth in chapter 5.

A similar problem is faced when attempting to make quantitative distinctions between the members of $\mathcal{I}_C$. One might, for instance, want to interpret different communicative actions as being caused by different amounts of utility differences. The fact that I_2 is willing to take a good deal of time out of her day to actively protest h might mean that she assigns a larger utility difference between h and $\neg h$ than I_3 does. Similarly, given that I_1 is willing to write to as well as call her representative she may also assign more utility difference between h and $\neg h$ than I_3 does. Such comparisons are especially arbitrary between I_1 and I_2. But really, all these conclusions about utility seem quite arbitrary. Again, the use of direct communications to infer attitude *strength* comes up against Arrow's worries about the basic subjectivity of individual assessments of cardinal utility. The ways in which individuals describe the degree to which they prefer some alternative over another is importantly subjective, and so it seems impossible to make comparisons

of cardinal utility from descriptions or actions that require complex background knowledge about each person in order to interpret the significance of their actions. For instance, I_1 might spend most of her day writing letters to her representatives on a wide array of issues about which she is only mildly interested, I_2 might spend all week protesting mostly for the thrill of it, and I_3 might only sign petitions when she is absolutely committed to a cause. In order to make sense of their communication behaviors, it is necessary to relativize them to each individual's behavior and psychological tendencies. That would seem to be too complicated a task for behaviors as complex as personal communications. I will discuss these general problems in greater depth in chapter 5.

Personal communications seem too complicated to allow for a meaningful assessment of public opinion from them. However, public opinion surveys are designed to regiment the communication of opinion information in just the way that is impossible for normal political communication. This may decrease the complexity of interpreting individual responses and allow for satisfactory representation of public opinion. However, next I will show that most modern opinion surveys measure only ordinal information about public opinion, and are therefore unsuitable for use with the weighing procedure. This is a problem that I will attempt to remedy in the next two chapters, providing a model of public opinion that is amenable to measurement by public opinion surveys.

3.3.3 Public opinion surveys

I will argue that public opinion *surveys* are natural methods of measuring public opinion, given the limitations of other measurement methods. But most current types of public opinion surveys only provide ordinal-level measurements, and are therefore inadequate to allow political representatives to abide by the *Duty of Political Responsiveness*. In the next two chapters I will lay the groundwork for better measures of public opinion, but I must first show that standard opinion surveys are indeed inadequate. Figure 3.2 is one example of a survey question asked in 2008 to a sample of Americans.[34] It uses relatively standard "Likert"-type survey items to produce ordinal scales of policy alternatives.[35] This type of scaling method is very common, not only in public opinion surveys, but also in social science research of many types.

The individual then assigns the subject of each question a position along the Likert scale. The ordinal interpretation of these responses is straightforward, except for the "Not Sure" item. There are two natural interpretations of "Not Sure": a position between Somewhat Favor and Somewhat Oppose and an expression of indifference (or lack of judgment). The first interpretation would call for a placement of the alternative in a middle position along the scale, whereas the second interpretation would require that the alternative not be placed along the scale at all. This is a scaling decision that relies upon there being an interpretation of "Not Sure"

[34]From the Los Angeles Times/Bloomberg News Poll # 2008-564: National Politics and the Economy, question #30

[35]I discuss the Likert scale in more detail in section 6.2.1.

> . . . Barack Obama is considering an economic stimulus package that would include tax cuts, an extension of unemployment benefits, and job creation through government spending on such things as rebuilding roads, bridges and schools. Such a plan would cost up to half a trillion dollars, and would increase the already record U.S. budget deficit to even higher levels. Generally speaking, do you favor or oppose a package like this that would stimulate the economy but cause the budget deficit to increase? (If Favor/Oppose) Are you strongly (in favor/opposed to) such a stimulus package, or only somewhat?
> 1: Strongly Favor
> 2: Somewhat Favor
> 3: Somewhat Oppose
> 4: Strongly Oppose
> 5: Not Sure

Figure 3.2: Likert-scale opinion question

that is shared by the respondents. For the purposes of demonstration, I will assume that no individuals selects "Not Sure" for any policy alternative.

It is important to notice that such scales must be ordinal, because no part of the questioning process establishes the distance between the various possible responses.[36] Nor does it provide any more information that allows for the meaningful comparison of utilities between individuals. In order to see this, consider another question that could have appeared on the same survey (though which I have invented), making use of the same response scale:

> Congressional Republicans have offered a stimulus plan that would only feature tax cuts. This plan would cost up to one quarter trillion dollars and increase the debt by that amount. Generally speaking, do you favor or oppose a package like this? (If Favor/Oppose) Are you strongly

[36] See Jamieson (2004) for a discussion of the ordinal nature of standard Likert scaling procedures.

(in favor/opposed to) such a stimulus package, or only somewhat?

... a third question:

> Some Congressional Republicans have opposed any stimulus plan, opting not to increase the debt at all from increased spending.

... and a fourth question:

> Some Congressional Democrats want to implement a larger stimulus package than what Barack Obama has offered. This would feature the same balance of stimulus measures, but would cost 1 trillion dollars (instead of ½ trillion). This would increase the national debt by that amount.

Each proposed policy can be thought of as an *alternative* under consideration. Apply the following labels to those policies:

x \$½ trillion comprehensive stimulus

y \$¼ trillion in tax cuts

z No stimulus

a \$ 1 trillion comprehensive stimulus

This allows for an assignment of *ordinal* utility values to each policy alternative. Consider some group of individuals $\mathcal{I}$, all of whom have rated every policy.[37]

[37] This means that none of them have responded with "NOT SURE."

Make numerical assignments to each type of response along an ordinal utility scale, with "Strongly Favor" being assigned '4', "Strongly Oppose" a 1, and so on. For instance, some individual I_1 may make the following responses for each question: (x, Strongly Favor) (y, Somewhat Favor) (z, Strongly Oppose) (a, Strongly Oppose), thus producing the set of assignments $\{(x,4)(y,3)(z,1)(a,1)\}$.

Assume the following assignments for the entire group:

I_1 $\{(x,4)(y,3)(z,1)(a,1)\}$

I_2 $\{(x,2)(y,1)(z,1)(a,4)\}$

I_3 $\{(x,2)(y,3)(z,4)(a,1)\}$

On the face of it, these measurements may actually give more information than simply *individual* ordinal scales; specifically, it seems to provide the *same* ordinal scale across all individuals. This is the extra information that is provided by the *labels* of the scale items. Given the coarseness of these labels (there are only four of them) they seem to provide a degree of *ordinal* interpersonal comparability. This makes it possible to say that someone who "Strongly Favors" some alternative favors it more than *someone else* who "Somewhat Favors" it. And even if this is denied, there seems to be no doubt that the difference between "Favor" and "Oppose" is interpersonally significant in this context. Someone who "Favors" an alternative certainly prefers it more than someone who claims to "Oppose" it. Whereas ordinal measures *without* interpersonal comparisons only permit statements of preference *within* individuals, ordinal measures with interpersonal comparisons permit comparisons between them. For instance, it can be said that I_1 rates y higher than I_2

rates x, even though both individuals rank those alternatives in their second-best ordinal positions. The labels used to make ordinal assignments are meaningful in this case, so it might be possible to use this information in a social choice function. However, it is not likely that all terms have the same meaning across individuals; this is likely only true for "Oppose" and "Support." Therefore, for the purposes of interpersonal comparability, it is possible to partition assignments into four classes, obtaining the following *intra*personal orders:

$I_1 \;\; x \succ y \succ z \sim a$

$I_2 \;\; a \succ x \sim y \sim z$

$I_3 \;\; y \sim z \succ x \sim a$

And the following *inter*personal order, where $I_n(F_i)$ is an individual (n) preference function and $[I_1(x), I_2(a), I_3(z)]$ means $I_1(x) \sim I_2(a) \sim I_3(z)$, when it is assumed (I think implausibly) that *all* scale positions are interpersonally comparable (and not just "Favor" and "Oppose"):

$$[I_1(x), I_2(a), I_3(z)] \succ [I_1(y), I_3(y)] \succ [I_2(x), I_3(x)] \succ [I_1(z), I_1(a), I_2(y), I_2(z), I_3(a)]$$

This is the most information that could possibly be gotten from an opinion survey of that structure, and because it does not just restrict significance to the "Favor" and "Oppose" groups of labels it actually makes more interpersonal comparisons of preference than are likely significant.

The interpretation of the scale that provides ordinal interpersonal comparisons is not actually more useful in forming a social preference function than the first set of *individual* scales. This is because no measure exists to provide the *differences* between utility assignments; the numerical utility assignments are only significant up to their ordinal information, and so any set of assignments that maintain this order is just as good as any other. The best social choice function available for ordinal scales is based upon the majoritarian voting method, which makes no use of the additional information provided by interpersonal comparisons of preferences. It should be stressed again that the above ordering is the best case scenario for interpreting these survey scales; I find it far more likely that only the difference between "Favor" and "Oppose" can be guaranteed to be equally significant between individuals. Similar considerations also doom any attempt to use scales with more positions, which is common for many Likert-type scales.[38] As the number of positions increases, it becomes less plausible that the significance of those steps can be compared between individuals. Therefore, Likert-type scales offer little more than ordinal-scale measureability and only rough ordinal interpersonal comparability. For now I think this is a satisfactory examination of the information derivable from Likert-type scales. I will return to this topic in section 6.2.1 when I consider whether there are any plausible assumptions that would permit *aggregates* of Likert scales to estimate cardinal measures of public opinion, even though *individual* scales are not cardinal. I will show that there is not much hope for this either.

[38]For instance, some Likert-scale items ask the respondent to select a response from a 5 or 10 point scale.

Public opinion surveys, and for that matter all standard methods of measuring public opinion for use by political officials, do not capture the amount of information necessary for officials to balance public opinion against what they believe is in the interests of the people. This problem with how public opinion is typically measured has received relatively little attention within political science.[39] However, the problem I will consider next has received a great deal of attention. Since the early years of public opinion measurement, there has been a concern that the public is too ignorant and uninterested in public policy for it to offer adequate opinions on many topics. Even before the widespread use of scientific methods to study public opinion, Lippmann (1922) and then Schumpeter (1950) rather famously argued against the populist view of democracy, stating that public opinion couldn't possibly serve as a coherent and intelligent basis for policy decisions. This old concern became a problem for survey methodology when Philip Converse's (1964) studies of early public opinion surveys showed that most people exhibited what appeared to be *random* opinions about most political issues. His findings began one of the most important debates in public opinion survey methodology, and it will be the other major problem (besides the need for cardinal measures) that any model of public opinion must overcome to be useful to political representatives.

[39]See Drasgow et. al. (2010) and Krosnick and Abelson (1994) for two discussions of this general topic.

3.4 Opinion instability

Beginning with Philip Converse's (1964) study of early American opinion surveys, there has been rather damning evidence that what is taken to be "public opinion" originates primarily from random thoughts of the respondents attenuated by features of the questioning context. Converse found that individuals' espoused opinions about various public policies fluctuated randomly over several years. Nearly the only thing that remained stable was the political party with which respondents identified. He concluded that most individuals simply do not have coherent attitudes (they have "non-attitudes") in most public policy domains.

The debate that followed Converse's findings indicated that although some of his results were due to measurement error, he was correct that most public attitudes are subject to widespread instability.[40] This finding greatly affected subsequent models of public opinion, including John Zaller's (1992) influential "Top-of-the-head" attitude model, which will serve as the basis for my treatment of policy attitudes. This model represents attitudes as sets of considerations that is open to each individual to use in deciding between the available alternatives. Although the *total* set of considerations available to make an opinion judgment is stable, the precise set of considerations used by the individual will vary between survey instances. Therefore, the opinion exhibited by an individual about an alternative may change between survey instances even though the attitude underlying those

[40]See Achen's (1975) famous response to Converse, arguing that Converse's "non-attitudes" were primarily the result of measurement error. Also see Converse and Markus' (1979) contemporaneous reply to Achen, arguing that response instability could not entirely be produced by measurement error.

expressed opinions remain the same. In the next chapter, I will develop such a stochastic model of attitudes that is compatible with expected utility theory, thus allowing the standard models of public opinion to interface with social choice theory.

I will first develop Converse's original problem in terms of a violation of the standard ordinal preference theory. Such a presentation of the problem is rarely used in political science, in favor of statistical formulations, but this is a rather natural representation of the problem that will lend itself to a social choice theoretic analysis. The preference model I will develop in the next chapter relies upon a notion of *preference instability* that is similar to the problem Converse originally found in responses to public policy questions.

One way in which preferences can fail to satisfy standard ordinal preference theory is by violating the condition of *transitivity*. Consider a set of alternatives F_1, such that the triple $x, y, z \in F_1$. To say that this triple is weakly transitive is to say that:

Weak Transitivity $x \succeq y \wedge y \succeq z \rightarrow x \succeq z$

Because (weak) transitivity is basic to most forms of rationality, it is quite difficult in normal situations to think of intransitive preference sets that are still intelligible. However, it becomes easier if it is imagined that an individual is asked to report her *pair-wise* preferences over some large set of alternatives. In that case, the fact that the pair is involved in a transitive triple may be hidden from the individual by the large number of pair-wise comparisons that she needs to make. Therefore, she can only rely upon whatever preferences she actually has (or at least

can access) that are relevant to each pair-wise comparison, and if her preferences over the alternatives within this triple are not especially well-formed then intransitive results seem plausible.

Consider an individual's preferences over the following alternatives, where all descriptions are changes to the *status quo*:

x Lowering taxes and increasing defense spending

y Lowering taxes and decreasing education spending

z Increasing taxes and increasing defense spending

Assume that she prefers x to y because she thinks that defense spending is, on the whole, one of the most important government functions. Later she claims to prefer y to z because she thinks that lower taxes are generally better than higher taxes. However, even later on she is presented with a pair-wise comparison between x and z. At this point she recognizes that given the importance of a strong defense increased defense funding may even warrant increasing taxes in order to avoid deficits. Therefore, she prefers z to x. Each pair-wise choice is made independent of the others, meaning that the resultant responses depend only upon the attitudes of the individual toward each of the alternatives, and not upon any constraints of rationality. If the individual is required to make these comparisons within a short period of time, then the logical relationships between the various preferences would likely be salient to her; therefore, it is likely that her choice between z and x would not be independent from the others. Rather, she is more likely to choose x over z

in order to conform to the requirements of transitivity. When transitivity does not provide a salient constraint on choice, individuals may make sets of pair-wise choices that when taken together do not conform to transitivity requirements. The case of intransitivity in the above example involves multidimensional choice, which is an obvious possible cause of such intransitivity. Many choices between public policy opinions involve a large number of *considerations* relevant to the choice. Therefore, any attitude theory that is thought to underlie public policy attitudes should treat this as the normal (and not a special) case. The attitude theory I present in the next chapter does this.

When the set of pair-wise comparisons is large enough, and therefore the individual is unaware of her previous expressed preferences, she may also exhibit preference *reversals*. Consider a variant of the example I give above. The individual could choose y over x if questioned again about her preference (she initially chose x over y). However, this means that that the following two preference assignments can be given to the same individual.

1. $x \succ y$

2. $y \succ x$.

This violates any standard preference model, all of which require the following condition to hold:

Antisymmetry $x \succ y \rightarrow \neg(y \succ x)$

I will call violations of this requirement *preference instability*.

As I described earlier, there is some indication that both unstable and intransitive preference expressions are common, and this has been perhaps best documented in the case of public policy preferences(Converse, 1964; Fishburn, 1991; Zaller and Feldman, 1992). Some have taken this to be an indication that populist views of democratic governance are doomed, because their core guide to action — the public "will" — simply does not exist.[41] Most arguments against this skepticism of public preferences have attempted to deny that instability or intransitivity exists on a wide scale.[42] However, I think that there is good reason to think that the preferences of the public are lacking in the ways that have been described.

Observations like those I give above have caused many to give up on the prospect of constructing anything resembling a social choice function out of the attitudes of the public. However, in the next chapter I will describe a theory of preferences that easily handles both intransitivity and instability. It will also show why the assumptions of the standard preference models are violated by many public policy preferences. This *stochastic* preference model accomplishes this by treating any given choice instance as merely a noisy *measurement* of a preference, and not a totally transparent display of it. This model also allows for the use of information that only exists for entire *sets* of choices, and not for individual choices. This will permit a description of a stochastic attitude theory in choice theoretic terms.

The purpose of this chapter has been to advance the following argument: Political representatives in democracies have duties to respect public opinion, and the

[41] In addition to Converse's writings, see DeCanio (2006).

[42] For instance, see the research spawned by Achen's (1975) findings that suggested that most preference instability is due to measurement error, and not unstable public policy preferences. Page and Shapiro (1992) provide perhaps the most important recent contribution to this view.

weighing conception of the balancing procedure is the best explication of this "respect." However, public opinion as it is standardly measured has two important problems. First, the ordinal opinion information typically produced by public opinion measures is not suitable for the cardinal information required for a weighing of public opinion against expected well-being. Secondly, widespread public policy preference instability makes standard preference models inadequate for representing it; my model of the weighing conception relies upon another sort of preference model. The model I present in chapters 4, 5, and 6 is meant to solve these problems and consequently show how political representatives can balance public opinion with their estimates of welfare.

Chapter 4: A Model of Individual Opinion

I will argue here that what is typically referred to as "public opinion" can be measured. However, a prerequisite for *measurement* of some phenomenon is its existence. Therefore, I must show that something matching the concept of "public opinion" that I introduced in the previous two chapters is likely instantiated in modern democracies. Additionally, it must exist in a form that lends itself to measurement. As I have already motioned and will explain further in this chapter, both of these have been disputed. One possible worry is that although public opinion exists, it may fail certain basic rationality requirements or be based upon incoherent or fallacious reasoning. These may render it unsuitable for inclusion within any public decision process. Consequently, its inclusion in the decision processes of political representatives would not fulfill *The Duty of Political Responsiveness* that I introduced in the last chapter.

Most modern models of public opinion follow in the tradition of L. L. Thurstone's psychological models of political attitudes.[1] However, there is wide disagree-

[1]Some may question whether the terms "opinion" and "attitude" should be treated interchangeably, as I will largely do from now on. 'Opinion' and 'attitude' certainly have different connotations in standard English as well as philosophy, but among those who systematically study public opinion, "public opinion" is largely thought to be the the collective *attitudes* of the public (or some aggregation of individual attitudes). Therefore, because I will not introduce a separate (non-attitude based) conception of public opinion, I will treat them largely equivalently.

ment about the relative contributions of the model's different components of public opinion. This is especially the case for the explanation of *instability* in individual political opinions, a phenomenon that I introduced in section 3.4. Respondents to public opinion polls often report different policy preferences when asked to make identical comparisons in different questioning instances. Some have argued that most of this preference instability is the result of genuine instability of attitudes themselves (Zaller, 1992; Converse, 2006), whereas others have argued that this instability is predominately the result of measurement error (Achen, 1975; Page and Shapiro, 1992). I will present a model consistent with either view.[2] Additionally, I introduce a formalization of the model that is also directly compatible with social choice theory, which is not the case for most models of public opinion. The first step in the presentation of this model will be to provide a method for the representation of unstable preferences. This will be required for my presentation of a public policy attitude model, and will also provide a clearer conception of preference instability.

4.1 Stochastic preferences

Preferences in social choice theory are typically represented by what are called *algebraic* preference models. However, there are alternative preference models that have received far less attention in the literature. One family of models, called *stochastic preferences*, represent preferences as *stochastic* relations.[3] These mod-

[2]My view does depend upon there being *some* underlying stable attitude. This might preclude Converse's (1964) most radical "non-attitude" conclusions about public opinion.

[3]Martin Peterson (2008), in chapter 4, gives a useful recent discussion of this family of preference models.

els have a number of advantages over algebraic ones, and are especially useful in the case of public opinion. Firstly, stochastic preferences can accommodate violations of transitivity and antisymmetry, both of which are prevalent in a number of theoretical contexts including public opinion about public policies (Converse, 1964; Fishburn, 1991; Zaller, 1992; Sen, 1995). I will pay particular attention to violations of *antisymmetry*, which can lead to violations of *algebraic* transitivity.[4]

A *stochastic preference* relation over alternatives x and y is defined as a pairwise relation $x \succ y$ with a certain probability P (I will call these values 'P-values') of that strict preference being exhibited in any given choice instance. This can be written as: $p(x \succ y) = P$. A stochastic preference with a P-value of $1/2$ represents indifference, whereas values away from $1/2$ represent increases in *stability*, with values of 0 and 1 representing perfect stability.[5] It follows that $p(x \succ y) = 1$ is a generalization of the algebraic strict preference relation $x \succ y$, because that algebraic relation is satisfied in every instance of choice. Similarly, $y \succ x$ can be generalized as $p(x \succ y) = 0$. Depending upon one's interpretation of P-values, they describe either the probability that the individual will make some choice in any given instance, the relative frequency of some choice over a set of choice instances, or a fact about some mental structure of the individual. And this, of course, is not an exhaustive list of possible interpretations. However, under all such interpretations, the individual is assumed to make a determinant choice in every choice instance. Additionally, although choices under stochastic preference models are not generally

[4]In section 3.4, I introduce "weak transitivity," which is an algebraic transitivity condition.
[5]Values away from 0 and 1 toward $1/2$ naturally represent increases of *instability*.

constrained to pair-wise comparisons between alternatives, the specific model that I will utilize is restricted to such binary choices.

Here I intend to find cardinal utility values using individual judgments, and so I will interpret P-values as the *frequencies* of judgments (or choices). Therefore, define a P-value for some relation $x \succ y$ as the frequency with which x is chosen over y ($n(x, y)$) divided by the number of instances of choice or, to use the language of scientific measurement, the number of *trials* (t). Thus the P-value can be interpreted as a representation of a set of choice behaviors or judgments:

$$P = \frac{n(x, y)}{t}$$

I will argue that choice behaviors in the context of public policy questions arise from underlying attitudes that individuals have about the alternatives, but P-values *themselves* cannot represent those attitudes. Rather, the *utility values* derived from P-values will represent (in some important sense) the attitudes individuals have toward the alternatives. I will argue that utility can represent the features of attitudes pertinent to the measurement of aggregate attitudes of a society. However, this will not become suitably clear until I have presented the entire model.

It can be rather easily seen that stochastic preferences will *not* satisfy some standard conditions of algebraic utility models. For instance, a common axiomatization of algebraic (strict) preference requires that the conditions of *anti-symmetry* and *transitivity* are satisfied (where x, y, z are alternatives):

Anti-symmetry: $x \succ y \implies \neg(y \succ x)$

Algebraic Transitivity: $x \succ y \wedge y \succ z \implies x \succ z$[6]

Choice behaviors of any individual with stochastic preferences with P-values other than 0 or 1 will obviously violate *Anti-symmetry*. This will also lead to violations of algebraic transitivity, even in cases where the *stochastic* preference has not violated a version of transitivity that applies to *patterns* of P-values. The satisfaction of the following condition called *strong stochastic transitivity* seems adequate, even though it allows for violations of algebraic transitivity:

Strong Stochastic Transitivity $p(x \succ z) \geq \max[p(x \succ y), p(y \succ z)]$[7]

This is only one of several possible versions of stochastic transitivity, though this "strong" version is implied by the utility model I will present in this chapter. In contrast to *algebraic* transitivity, this is a condition on the *set* of all comparison instances of a triple of alternatives, and it therefore can withstand the violations of anti-symmetry that I have called preference "instability". *Strong stochastic transitivity* is a far more difficult condition to violate than algebraic transitivity, which can be violated by intransitivity in any single set of three choice instances involving the triple. The ability of stochastic preferences to withstand and, more importantly, measure preference instability makes them particularly useful in domains where preference instability is common and linked to utility differences between alternatives. This is the use I wish to make of such stochastic preference relations. In section

[6]Note that this condition is stronger than *weak transitivity*, which I introduced in section 3.4. This condition is required for *strict* rather than *weak* preference relations. I restrict my model to strict preferences, because the specific model I use is based upon binary choices, which are naturally represented by strict preference relations. However, not much of deep conceptual significance hangs on this modeling choice.

[7]This means that $p(x \succ z)$ must be greater than the larger of $p(x \succ y)$ and $p(y \succ z)$.

4.4.2 I will provide a means to find individual utility scales that can represent attitudes from a set of individual stochastic preferences.[8] However, first I must give an attitude theory that will serve as the basis of both individual utility *scaling* and *interpersonal comparisons* of utility.

4.2 Public policy attitudes

In this section, I will give a formal analysis of those public policy attitudes that are relevant to *The Duty of Political Responsiveness*. However, it must first be made conceptually clearer what those attitudes are like. From now on, I will denote all such attitudes as "*policy attitudes*." I will be concerned with two primary elements of policy attitudes through two distinct (though related) attitude models: their possible rational basis (shown in figure 4.2) and the way in which they can be modeled using cardinal utility representations (figure 4.4). These models describe policy attitudes in different ways, thus making them capable of answering different questions. A utility scale only represents the relationship between policy alternatives in terms of the degree to which individuals affirm each alternative, and so does not necessarily resemble the cognitive structure of that attitude. Similarly, the other model will demonstrate how stochastic preferences might be based in a rational process, though I leave open the possibility that some public policy attitudes sometimes arise from non-rational processes. And again, this rational process need not match all (or even most) features of the cognitive process that instantiates the

[8] I mean 'utility' in its purely formal sense, as a numerical representation of preference. Although I will make this clear later, the reader should not import any other non-formal conception of 'utility' at this early stage of the presentation of my model.

rational attitude structure. Therefore, the purpose of the rational model is not necessarily descriptive, though it does provide some descriptive constraints on the type of psychological structure that might be counted as a policy attitude. Whereas the purpose of the utility model is to represent a very restricted element of the attitude, the rational attitude model is primarily *justificatory* in nature; it demonstrates how policy responses might be unstable while the *attitudes* that underlie those responses are nonetheless stable and rationally formed.

Given the background theory of popular sovereignty that I have advocated, some of the features of policy attitudes are likely already clear. But at this point I must make explicit the necessary properties of those attitudes. Firstly, a policy attitude should be framed from the perspective of a neutral social planner, who makes a policy choice based upon what is in the common good rather than what is in the self-directed interests of the individual who has the attitude. Therefore, a policy attitude is not simply an expression of self-interest or desire, but is an expression of the *judgment* of the individual about which policy should be selected.[9] Public policy attitudes are modeled after the *general will* that I discussed in chapter 2, which I claimed underpinned the *Duty of Political Responsiveness*. As a model, the general will also serves as a normative ideal for individual political attitudes that direct governance toward the common good. Additionally, the general will provides an ideal to political representatives about how they might aid their constituents in making decisions that are aimed at the common good. But although the existence of

[9]I am leaving strategic considerations aside; what a constituent thinks her representative should do from a *strategic* standpoint is different from her policy attitudes, although policy attitudes will often dictate the strategy that should be used to achieve them.

the ideal of common-good directedness is a necessary condition of policy attitudes, it is *not* a necessary condition that all policy attitudes live up to this ideal. Popular sovereignty, which is at the basis of the general will, requires that the judgments of individuals be respected by representatives regardless of the manner in which individuals arrived at those judgments. Political representatives do not generally have a right to ignore policy attitudes just because they judge that those attitudes were arrived at illegitimately. This is one feature that distinguishes the view of popular sovereignty that I have defended from the deliberative democratic views that I have rejected. I discuss ways in which representatives might weigh opinion differently based upon the reasonableness of the opinion formation process, while still abiding by the *Duty of Political Responsiveness*, in section 6.1.

The second necessary condition on policy attitudes is that they must be derived from *judgments* about political policies, and these judgments must be explicitly endorsed by the individual. This excludes so called "*implicit* attitude" measures that have recently shown promise in predicting behavior by deriving attitudes from behaviors and physiological reactions seemingly unrelated to those attitudes.[10] Such implicit attitudes are clearly not *endorsed* by the individual in the way that I require. One of the best examples of implicit attitudes are those about race. One rather robust finding shows that Americans are more likely to associate Caucasian faces with positive words and images, and faces of African Americans with negative words and images. This is the case even when individuals do not themselves report feeling biased against African Americans, and occurs across racial groups (Nosek et al.,

[10]See Petty et al. (2012).

2002).

One may think that if representatives should be in the business of understanding the public policy attitudes of their constituents then they should also seek to understand *implicit* attitudes of their constituents as well. For example, it might be possible to give individuals implicit attitude tests to determine whether they have more negative associations with "higher unemployment" or with "higher prices." If an individual has more negative associations with "higher prices" then it may be reasonable to say that the following preference relation should be ascribed to that individual's attitude set:

$$\text{(Higher unemployment)} \succ \text{(Higher prices)}$$

But given the unconscious character of implicit attitudes, that individual may consistently claim to prefer higher unemployment to higher prices, thus *endorsing* the preference of:

$$\text{(Higher prices)} \succ \text{(Higher unemployment)}$$

In this way, implicit attitudes are contrary to the purpose of public opinion in democratic *governance*; if public judgments (instantiated in public opinion) are to play any role in self-governance, then the *explicit* judgments of the people must actually be measured and respected by political representatives.

The sorts of attitudes I am concerned with must also be distinguished from

emotional responses to policies, which are often thought of as attitudes. People who have policy attitudes that support tax decreases will also tend to experience negative emotional affect when told about likely tax increases, and will tend to vote for political candidates who support tax decreases. However, such "attitudes" will often come apart conceptually. Even *socialists*, who support high taxes, will often experience negative affect when told that *they* must pay higher taxes, and may even tend to vote against candidates who admit that they intend to raise taxes. Much of the work by psychologists studying attitudes has been devoted to finding the interconnections between attitudes, emotion, and behavior, but that is not what I will be concerned with here.[11] Rather, I will present a model that is limited to the link between the structure of attitudes within an individual's attitude set and the choices by that individual over a set of possible policy alternatives.

The complete structure of policy attitudes must be *inferred* from direct, explicitly endorsed choices among a feasible set of policy alternatives. I will find the relative strength of a political attitude toward an alternative by way of the relative frequency with which that alternative is chosen over other alternatives within the set of relevant policy alternatives. Therefore, attitude strength is inferred experimentally from the set of choices by an individual. Although these attitudes are the result of directly endorsed choices by the individual, the relative strengths of attitudes are *not* endorsed. In this way, the structure of public policy attitudes can only be found through a measurement process that is similar to the measurement of other psychological phenomena. This requires some explanation, because it may

[11] See Eagly and Chaiken, 1993.

seem that this suffers from the same problems of the implicit attitudes I mentioned earlier. And this may be the main problem with any measurement-based process for understanding the "will" of the people; whereas with direct communication and voting people are fully responsible for the opinion-relevant actions they take, in the case of public opinion surveys the public opinion must be interpreted by way of some sort of scientific measurement process. Furthermore, when only a *sample* of the people are included in the survey, the opinion of the entire population must be inferred from the opinion of that sample. For now, set aside the issues relating to sampling methods, as I will address these in section 6.2.1.

It may be that individual expressions of opinion should only be taken at face value, and that to infer *anything* from them is to not fully *respect* them. What should be made of such a claim? I think that the core of this objection is as follows: to *infer* some underlying attitude through attitude measurement, when some features of the measured attitude are not explicitly endorsed by the individual, subordinates the judgment of the measured individual to that of the person doing the measurement. There is some plausibility to this. For instance, when a scientist or a doctor takes measurements, the subject and the patient (respectively) are indeed in a subordinated position. However, this measurement paradigm needn't have such a subordinating effect. There are non-subordinating situations that can also be described as a form of measurement. For instance, the interpretation of communications might be thought of as a form of measurement. When I walk down the street and someone gives me a look of annoyance, this is certainly an *expression*, but it has little value unless I take it to be signaling (or at least linked to) some *attitude(s)*.

Specifically, I assume that there exists some attitude (not a very nice one) that has *me* as its object. Likewise, many linguistic expressions are notably expressions of some *attitude(s)*, and often require some inference to determine which attitudes they express. Although the statistical inferences that are used in standard public opinion measures are more *systematic* than the sort of measurement and inferences required for the understanding of direct communication, I do not think that it is different in *kind* from other forms of communication. As I hinted at in section 3.3, common methods of political communication are often difficult to systematically interpret in ways that are necessary for their use in political decisions. I take the main purpose of public opinion surveys *not* to be the treatment of individuals as experimental subjects, but rather to aid in political communication. Opinion surveys allow constituents to communicate their attitudes to their political representatives in ways that facilitate the use of those communications in political decisions.

The requirement of communicative regimentation is one of the most substantial ways in which the public opinion measurement that I will present differs from direct political communication. I require that the full structure of an attitude *set*, as represented by a utility scale, is derived from behavioral data about *pair-wise* choices between policy alternatives. Therefore, the pair-wise choices between policy alternatives is directly communicated by each individual, but a further set of information is *inferred* from the patterns of identical pair-wise choices within different *choice situations*. This process will require a great deal of explication, but the general idea is this: the features of the choice situations in which some individual makes choices between *identical* alternatives can differ substantially. For instance, varied

question wordings may lead an individual to think about a choice between two alternatives in slightly different ways. However, these differences are *not* relevant to political decisions; only the attitudes between the alternatives *themselves* are. I will introduce a method to derive information about the attitudes from choices.

Individuals are not capable of accurately communicating the cardinal structure of their attitudes; this is a major reason why the indirect method I present is necessary. There are two major considerations that support such a claim. They are related in important ways. The first arises from the economics and social choice literature, which shows the insignificance of cardinality in individual reports of utility. In fact, Arrow famously claims in *Social Choice and Individual Values* that such measurements have no place in social choice theory at all.[12] This is because such reports are wholly *subjective*, and there exists no straightforward way (if any way at all) to make *comparisons* of subjective cardinal utility reports between different individuals. The second set of considerations against the direct measurement of cardinal utility can be found primarily in the psychological literature, and is related to the problem of self-knowledge. Individuals seem particularly bad at reporting their attitudes toward various objects, be they public policies or not.[13] The choices people make are often greatly affected by features of the choice situation that are entirely unrelated to the actual choice at hand. This is bad enough for *direct* choices or rankings of alternatives, but matters become even worse when individuals are

[12]In fact, in the first sentence of chapter 2, Arrow states "The viewpoint will be taken here that interpersonal comparison of utilities has no meaning and, in fact, that there is no meaning relevant to welfare comparisons in the measurability of individual utility" (pg. 9).

[13]See Carruthers (2011) for a broader discussion of the problem of self-knowledge in psychology that extends to a large class of propositional attitudes, from beliefs to emotions.

asked to report the *cardinal* structure of such rankings. It is unreasonable to expect reported cardinal attitudes to be significant except in a very crude way.

A simple example will show how the problems of interpersonal comparability of cardinal utility and self-knowledge are related. Consider a process by which individuals report an entire set of attitudes toward a set of alternatives F_i. The simplest method, and one that does not impose an outside interpretive structure on *responses*, is a free assignment of real numbers to each alternative. However, it is quite obvious that such a method would make interpersonal comparisons of cardinal responses difficult. For instance, it is unclear what significance to place on the fact that one individual's smallest utility difference is 1 whereas another individual's smallest utility difference is 100. Therefore, I will consider a slightly more complex method in which the basic *limits* of utility assignments are constrained though utility *differences* within those limits are free to vary. Individuals may be told that they can assign alternatives any numerical value within the interval $[-10, 10]$, and that this is to stand for the relative strength of their preference for those alternatives. Instructions might also require that positive values only be assigned to alternatives with "positive" valence, and negative values for those with "negative" valence, but initially I will assume that this use of the scale is left to vary between individuals. Figure 4.1 shows an example set of assignments by all three individuals in $\mathcal{I}$ over alternatives $x, y, z \in F_i$.

There may be little difficulty in forming orders from these responses, and in fact there is even a *social* order that can be found from the set of individual orders

φ_2 \ φ_1	x	y	z
I_1	-8	7	10
I_2	2	4	0
I_3	-5	0	-2

Figure 4.1: Utility Assignments by $\mathcal{I}$ over F_i, constrained to the interval $[-10, 10]$

using the Condorcet majoritarian method ($y \succ z \succ x$). But given the lack of details about the ways in which individuals made utility assignments, it is impossible to know the relative significance of utility assignments *between* individuals. Individuals may use negative and positive numbers in different ways, and nothing about this method allows for a common interpretation of all individual responses, even though the *numerical* representations of these attitudes imply that such an interpretation can be made. In fact, not even the significance of utility *differences* can be inferred from such austere data. Compare the utility differences for I_1 and I_2. Do the greater differences in assignments by individual I_1 signify greater strengths of attitudes than I_2? Or are these individuals' response *styles* merely different? It is not clear without further information about the *source* of those utility assignments; a definitive answer to the question of significance requires some knowledge about whether the source of these differences is located in the attitudes themselves or in some other aspect of the individuals' psychologies. What is required is some basis for the *comparison* of these responses, such that the underlying cardinal structure of *attitudes* can be isolated from other phenomena that affect the *numerical* assignments.

From the above discussion, one might think that the cardinal structure of individual attitudes might be directly elicited if only some further method is found

to *calibrate* individual responses with each other, or if the elicitation is structured so as to force individuals to use utility differences in similar ways. However, individuals also seem to lack direct access to the cardinal structure of their *stable* attitudes, even for *emotional* attitudes, which *seem* to have clear cardinal structure.[14] I think matters are even worse for public policy attitudes. In order to report cardinal information about one's attitudes, it is necessary to have abstract knowledge of how one's preferences for various alternatives are related to one another. It is unclear how one could even determine such a structure except by considering individual comparisons of utility differences. For instance, an individual might be able to determine that some policy x being adopted instead of another policy y is more important for her than a third policy z being adopted over some fourth policy a. Thus, the following statement would be true of her attitudes:

$$u(x) - u(y) > u(z) - u(a)$$

If this individual could develop an order for all utility differences in her attitude set, this would enable her to claim self-knowledge of the cardinality of her attitudes. In fact, this is the most limited amount of information about an attitude that could still be counted as cardinal information. But it is highly unlikely that any individual could report such a structure without exhibiting either *instability* or *non-transitivity* in those differences. The most plausible method by which that information might be constructed is through a measurement method similar to what

[14] See Carruthers (2011), chapter 5.

I will propose, one that constructs information about attitudes through a series of identical comparisons. These comparisons permit instability information to be utilized in the construction of the utility scale. This method requires that certain assumptions be made about the basic structure of public policy attitudes.

I will present a schematic version of a very general type of stochastic attitude model that relies upon only very general features common to several competing attitude theories.[15] However, there is wide disagreement about the relative contributions to public opinion of the model's various components, especially about the relative contributions of error and stable policy attitudes. These models all follow in the tradition of Thurstone's (1959) "Law of Comparative Judgment." My model will be based upon Thurstone's basic attitude model. Additionally, I present the model in a form that is directly compatible with social choice theory, which is not the case for standard public opinion models.

First, I will present an attitude model that shows the possible rational basis of policy opinions. This attitude theory explains how unstable expressed preferences might arise from a reasoned attitude formation process, but this need not require that all individuals form policy attitudes in this way for those attitudes to demand the responsiveness of political representatives. As I claimed in chapters 2 and 3, individuals who are citizens of a democracy have the right to have their policy opinions respected by their political representatives, although this does not extend to having the reasons they used to formulate their attitudes respected if those reasons are

[15]For instance, see Achen (1975); Zaller and Feldman (1992); van der Veld and Saris (2004); Converse (2006).

not publicly acceptable. Consequently, the requirement of respect for opinion does not permit political representatives to "launder" the policy attitudes of individuals, eliminating those attitudes that are not based upon publicly acceptable reasons, beyond possibly excluding *unjust* policy alternatives.[16] However, this does not preclude the "laundering" of the *reasons* behind those preferences when representatives consider what reasons should be included in political deliberation.

Assume that any individual preference (from now on I mean *stochastic* preference) between two alternatives is determined by some weighing of *considerations* for and against each alternative. This, of course, presents a series of problems related to how considerations can be individuated and how the weighing procedure should be thought of (or whether this should be thought of as a weighing procedure at all). I will put these problems to the side, because any theory of reasoning will face them, not just the attitude model I present here. However, if it can be assumed that different considerations can be *categorized* based upon the types of effects they have on the resultant attitudes, then only the *effects* a given set of considerations has on the attitude will be necessary to determine the resultant choices. Therefore, one might also dispense entirely with the use of considerations in a model of attitudes without any serious harm to how the theory predicts choice. Such a simplification strategy will become useful later when I discuss the representation of attitudes using utility theory, but such a simpler model does not explain as well the possible rational basis of policy opinions.

[16]See Goodin (1995) on the various ways that individual preferences might be restricted or adjusted in order to eliminate preferences that exhibit unjust or otherwise unacceptable (for instance racist) preferences.

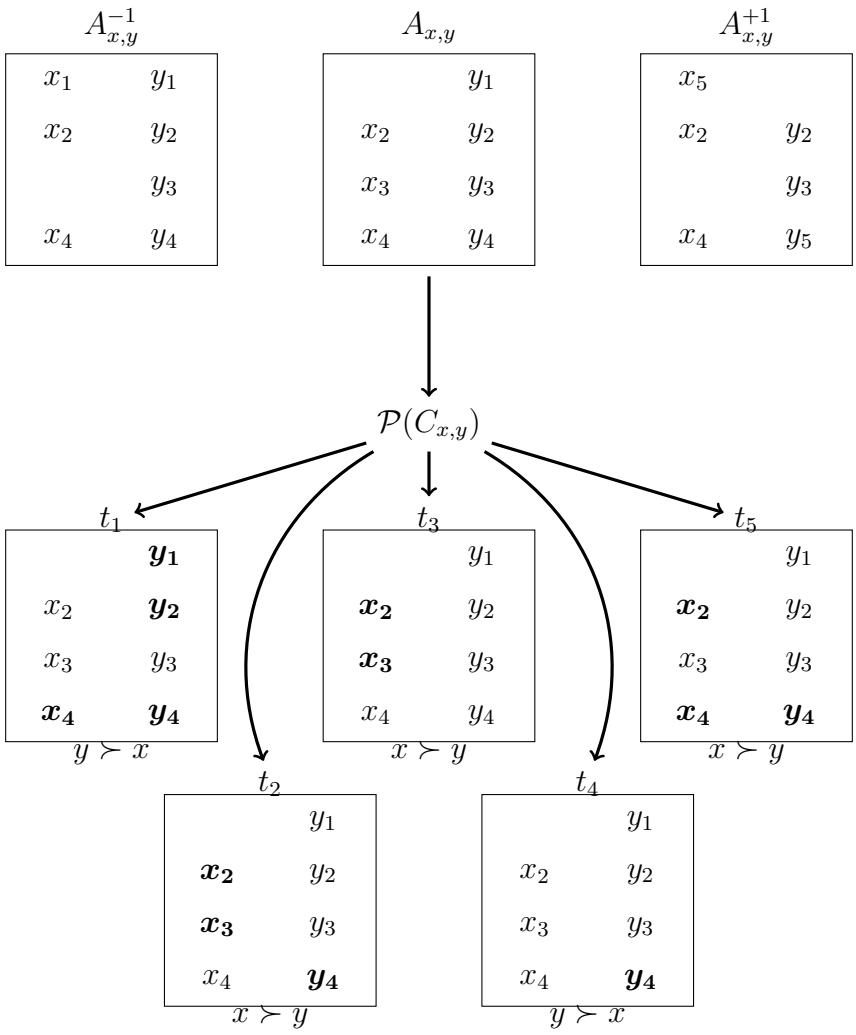

Figure 4.2: Attitude model

The full attitude model is shown in figure 4.2. This represents the attitude of an individual toward a *pair* of alternatives $x, y \in F_i$. The resultant *pair-wise* attitude, $A_{x,y}$ (top-middle of the figure), is the psychological structure that determines choice patterns between those *two alternatives*. This *pair-wise* method of analyzing attitudes will likely seem unnatural to some who think of an attitude as being about some particular (single) alternative. I will give such an analysis of attitudes, but it will require the development of a utility theory first; the attitude an individual

has towards an alternative will be represented by the utility value she assigns to that alternative, relative to some set of alternatives under consideration F_i.[17] But at that point I will eliminate all reference to *considerations* from the model, in favor of a simpler analysis. However, so long as considerations remain in the model — and they will be necessary to justify the simpler version of the model — a pair-wise definition of attitudes is required. Why this is will become clear shortly.

First consider the case of a *static* attitude that is not susceptible to any attitude change. The static attitude $A_{x,y}$ has two contents: a set of relevant considerations $C_{x,y} = \{x_2, x_3, x_4, y_1, y_2, y_3, y_4\}$, and a probability distribution $\mathcal{P}(C_{x,y})$ over the combinations of considerations in $C_{x,y}$ that are used in any instance of choice.[18] The total set of considerations in $C_{x,y}$ is shown in the box below the label '$A_{x,y}$'. The probability distribution ($\mathcal{P}(C_{x,y})$) represents the *salience* or importance of each combination of considerations to the person's deliberations over the pair of alternatives.[19] Finally, assume that every alternative favors one or the other alternative and that each consideration is equally favorable to that alternative as any other. In other words, no consideration has more "force" than any other. Therefore, the differential weight given to different considerations must be represented either

[17] As I will show in section 5.1, this will eventually need to be broadened to all alternatives in the *choice context* (D_i), which will allow for *interpersonal* comparisons of utility values.

[18] Given that the number of combinations increases exponentially with the number of considerations, this distribution will be rather complex and difficult to calculate in sets much larger than a few considerations. Therefore, it may be more intuitive to replace this element of the model with a probability function over *individual* considerations, though that function would not fully determine the resultant choice pattern ($P - value$) of the individuals for that alternative pair.

[19] This model is similar in certain respects to a recent reason-based preference model presented by Dietrich and List (2013). Their model features "reasons" that modify the resultant individual preference based upon which of those reasons are salient in any given choice context. However, their model lacks an endogenous explanation of how certain reasons might be more often salient than others. Because this is an important aspect of public policy attitudes, I include this in my own model.

in the probability distribution or in the number of considerations favoring a given alternative. The basis of salience in the basic structure of the attitude is important, because it is this structure that I will attempt to measure through a utility theory. Salience, therefore, is a *normatively* significant feature of the attitude, and not merely the result of aspects of human psychology unrelated to the individual's will. This attitude theory posits that there is an important stochastic nature to attitudes; the same attitude is not always expressed in identical ways, even though it may have a stable structure. And it is that entire structure that is relevant to public opinion and the sovereign will of the people.

The choice an individual makes between two alternatives x and y in some particular instance of choice is determined by two steps. The first is the set of salient considerations, given particular features of the questioning environment, how the comparison is made, the individual's state of mind, and so on. The second step is the random *measurement error* that affects the choice without affecting how the individual thinks about the question. Some features of the measurement instrument (if, for instance, it uses a Likert-type scale that has options from 0-10 or -5 to 5) may affect the sorts of responses an individual gives, even though it does not change the sort of considerations the individual uses. This does not affect the attitude itself and thus is not an important feature of the attitude model. Measurement error is a problem for any type of measurement, and because it is not a result of features of the stable attitude its features are not relevant for policy decisions. Therefore, as a conceptual matter, I will largely ignore it. I will primarily use it as a random factor that decides the choice when there are an equal number of

considerations favoring each alternative, but this is just for convenience, and could be easily changed without harm to my view of policy attitudes. In fact, I recognize that this use of measurement error is a bit misleading. Ideally I would use it as a factor that randomly alters the policy choice from any set of considerations, with the probability of that occurring as a function of the number of considerations that would need to be changed in order to alter the choice. However, this would yield an unwieldy model. In part this is an artifact of the particular way I am structuring the full version of the attitude model, and there will be a much more natural way to model measurement error once I give the simplified attitude model.

The full attitude model allows for attitude change through the addition and subtraction of considerations, though I only show the possible results of such attitude change. $A^{-1}_{x,y}$ is an initial set of considerations with $C^{-1}_{x,y} = \{x_1, x_2, x_4, y_1, y_2, y_3, y_4\}$. This set is subject to change through whatever orderly process exists to make such change. So, for instance, such a process allows for x_1 to be dropped from $C^{-1}_{x,y}$ and x_3 added to $C_{x,y}$. I will not discuss how attitude change might function, but the key for my purposes is that it can produce *stable* attitudes. Even though such a stable attitude may change over time, at any given time there is some set of considerations that constitutes that attitude.

The stochastic element of the model is represented by a probability distribution over the set of considerations, which determines the probability that any ratio of considerations in favor of x or y is selected in some given instance of choice. For each trial (t_i) a set of considerations is chosen from the set of all considerations in $A_{x,y}$ in accordance with the probability distribution $\mathcal{P}(C_{x,y})$ over $C_{x,y}$. The choice

of alternatives in that instance is determined by the set of considerations contained within that selection of alternatives. In figure 4.2, the squares labeled 't_1 – t_5 show consideration selections from $C_{x,y}$ for each trial in bold font. But in this depiction of the model, the probability distribution $\mathcal{P}(C_{x,y})$ is not detailed, mostly because it would be very complex for a set of seven considerations if all combinations of considerations are included in the distribution. However, a simpler example can illustrate how such a probability distribution can lead to policy responses. Consider a very restricted $C'_{x,y} = \{x_1, y_1\}$, contained in $A'_{x,y}$. In this case, $\mathcal{P}(C_{x,y})$ can be given with only three probability values. An example of such a probability distribution is shown in figure 4.3.

Salient Considerations	x_1	x_1, y_1	y_1
Probability	.4	.4	.2

Figure 4.3: Probability distribution

Assume that error related to measurement only affects individual choice in the case of balanced considerations (x_1, y_1), and that it results in the selection of x or y at chance level. $\mathcal{P}(C'_{x,y})$ would result in a selection of x over y with a probability of 0.6 (and y over x with a probability of 0.4).[20] Using the $P-value$ notation of stochastic preferences: $p(x \succ y) = 0.6$. The rational attitude model, with its reliance upon the salience of various considerations, offers a number of possible interpretations for any given probability distribution over considerations. For instance, the relative

[20] x is chosen over y with $p = 0.6$, because x_1 has a probability of 0.4 of being the only consideration used to make the choice, and the probability that both x_1 and y_1 are used has a probability of 0.4. If x_1 is the only consideration, then x is chosen straight-away. If x_1y_1 is used then x is chosen with a probability of $1/2$. $\cdot 4/2 = 0.2$ and $0.4 + 0.2 = 0.6$.

salience of a consideration $\mathcal{P}(C'_{x,y})$ might be thought of as the relative importance the individual places on the consideration. This permits an easy way to understand what aspect of an attitude can be stable, even when different expressions of that same attitude might bring about different results. The considerations involved in choice and the relative salience of those considerations are the stable aspect of an attitude. Of course, the term 'salience' carries a descriptive connotation in psychology. Some considerations might simply be more likely to be thought of, perhaps because they are consistent with some sort of cognitive heuristic or more easily understood and remembered.[21] This is a problem for any theory of rational choice, and so I do not think I need to take it on specifically. It is not my purpose here to show that people are exceptionally rational in all of their policy attitudes; I have simply intended to show that individuals could have stable, rationally-based policy attitudes that are nonetheless susceptible to unstable expressions in policy opinion measurements.

Next, I will show how sets of attitudes might be represented along a cardinal utility scale. Not all of the details of the model I provide in figure 4.2 are important for the utility analysis, and so I will drop much of this detail at this point. I will also present a simplified attitude model (in figure 4.4) that will make clearer how the cardinal properties of policy attitudes are formed. It must be remembered that all uses here of 'utility' are purely formal, such that utility values are simply numerical representations of attitudes; the subject of analysis (policy attitudes) does not differ in the simpler model.

First, assume that there exists a set of *static* attitudes about some alterna-

[21] For an introduction to such cognitive phenomena see Kahneman (2011).

tive φ_1, and all other alternatives in F_i, where $\varphi_1 \in F_i$. φ^* can be any other alternative in F_i. As before, each alternative *pair* in F_i is the subject of a distinct attitude. Therefore, if $F_i = \{x, y, z, a, b, c\}$, then there are five attitudes associated with alternative x $(A_{x,y}, A_{x,z}, A_{x,a}, A_{x,b}, A_{x,c})$. The utility analysis will permit the representation of the entire set of attitudes by a single utility assignment, which has cardinal significance along a utility scale containing the other alternatives in F_i. This simpler representation will allow for a deeper analysis of individual attitudes and how such a model may permit *interpersonal comparisons* of attitudes.

A few key assumptions will allow for a straightforward method by which $P - values$ can be said to originate from the utility assignments to alternatives. First, any effects due to transient features of the choice situation (framing effects, contextual variability, etc.) and measurement error can be captured by a single element of the model T, because I will assume that the probability distributions of their effects are all normally distributed and independent; such assumptions permit the conclusion that the combination of those effects are also normally distributed.[22] Note that both the stochastic effects from the probability distribution over the set of considerations $(\mathcal{P}(C))$ as well as measurement error in a specific choice context are modeled by this single element of error T. The stable element of an individual's attitude toward φ_1 can then be represented as a utility value, $u(\varphi_1)$. This "true" attitude toward φ_1 is the utility scale value derived from a comparison of φ_1 with all other alternatives within the set F_i. Given the three elements $(u(\varphi_1), u(\varphi^*),$

[22]Thurstone makes this assumption in the simplest formulation of his attitude model ("*Case 5*"), though it is not a necessary assumption. More complex formulations allow this assumption to be dropped.

and T), the probability $p(\varphi_1 \succ \varphi^*)$ that an alternative φ_1 will be chosen over some other alternative $\varphi^* \in F_i$ in any given choice instance is determined by the attitude model in figure 4.4.

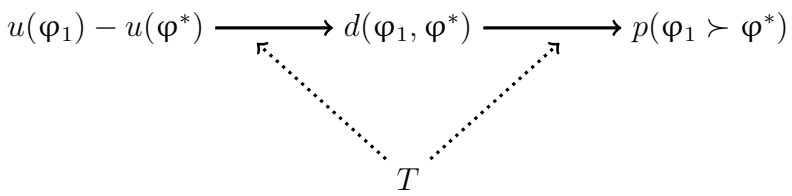

Figure 4.4: The attitude toward alternative φ_1

This simplified attitude model assumes that the individual is forced to make a set of choices between two alternatives in pair-wise comparisons, which will often result in preference instability. When the contribution of $u(\varphi_1) - u(\varphi^*)$ on $p(\varphi_1 \succ \varphi^*)$ is small, or that of T is large, then $p(\varphi_1 \succ \varphi^*)$ will be close to ½. T can be thought of as acting in two steps of the model. It first acts on both $u(\varphi_1)$ and $u(\varphi^*)$ in the determination of $d(\varphi_1, \varphi^*)$, which is the measured utility difference between those alternatives. This accounts for transient effects that arise from some particular *comparison context* of any two alternatives and the effect of that comparison context on the probability distribution over considerations ($\mathcal{P}(C)$) from figure 4.2. This transient aspect of the attitude model arises for all instances in which those alternatives are compared, but it is distinct from the difference between the utility assignments to those alternatives when those utility assignments are calculated through *comparisons with all alternatives*. Next, T acts on the expression of $d(\varphi_1, \varphi^*)$ in $p(\varphi_1 \succ \varphi^*)$. This represents both response error and transient effects from factors within the particular *choice instance*. The model allows for the isola-

tion of T from the stable component of the individual's attitude $u(\varphi_1)$, such that $u(\varphi_1)$ can also be isolated as the portion of the attitude that is relevant to social decisions.

In order for this model to be relevant to public opinion measurement, it must be plausible that a numerical representation of an attitude — a *utility* value — is able to characterize the features of attitudes that are relevant to social decisions of public policy. It is possible that the stable portion of the attitude *represented by* $u(\varphi_1)$ is relevant to the social decision, even though the *numerical form* of that attitude is not. In order for this representation to be relevant it must be meaningful to place the strengths of an individual's preferences, at least within each particular policy domain (or *choice context*), on a single cardinal scale. However, such a representation is a significant departure from the attitude theory typically given by the attitudes literature in psychology. I must be more explicit then about the way in which some utility scale can be said to represent a set of attitudes. In my view, the utility scale represents the degree to which an attitude supports the choice of the target alternative of that attitude relative to the degree to which the attitudes of other alternatives support their target alternatives. Given that there is no objective standard with which to compare attitudes, this model must rely upon *relative* judgments. In section 5.2, I will present a means of interpreting each individual set of responses by making use of certain descriptive statistics of each individual scale, based upon psychological characteristics of the individual. This strategy is intended to allow for a proper interpretation of each individual attitude by making use of certain facts about each individual's attitude structure.

Before I move on to a consideration of how individual utility scales might be constructed, I will consider a problem that arises for most attitude theories. It may be the case that the precise identity of attitudes measured in any given choice instance cannot be assured, because it can never be guaranteed that two different choice instances actually involve *identical* attitudes. This is similar to the problem of identifying the (proper) *referent* of an attitude. Differences in question wording is one of the simplest examples of how different choice instances are thought to nonetheless be instances of the same attitude. For instance, if one wants to determine whether an individual prefers more money spent on the arts (keeping all other funding stable) or more money spent on space exploration (again keeping all other funding stable), it is likely that the exact question wording could have an effect on the set of considerations an individual uses in her response (t_ns from figure 4.2). Therefore, when measuring an attitude itself and attempting to avoid systematic biases due to particulars of question contexts, one may use several different wordings and question frames. However, it is not clear that asking an individual whether she would rather support "the arts" than "space exploration," is similar enough to "public funding of visual art, music, and other media" and "the public development of space travel technology" to trigger identical underlying *attitudes*. These two comparisons may produce fundamentally different stable considerations, and therefore might best be modeled with two different pairs of attitudes. This produces a significant worry that the theory of attitude measurement I have provided also requires a substantive theory of attitude identification. Although I recognize that this is an interesting problem, it is not a problem unique to my theory of attitudes,

but rather is something at the center of many contemporary debates among survey methodology theorists.[23]. Therefore, I will set this problem aside for the purposes of this dissertation. I will assume that differences in semantic interpretation of a question can be properly contained within the probability distribution over possible consideration sets.

In this section, I presented a model of attitudes (in figure 4.2) that I think shows how reason-based policy attitudes might produce unstable (stochastic) policy preferences. Additionally, I have shown how these attitudes can be represented in terms of differences of utility (in figure 4.4). In the next section, using this utility attitude representation, I will show how individual utility assignments can be derived from sets of stochastic preferences. This will then allow for a deeper analysis of individual attitude sets, and eventually the significant comparison of those attitude sets across individuals, resulting in an aggregate cardinal measure of individual attitudes (i.e. a cardinal measure of public opinion).

4.3 Individual utility

There are a number of stochastic utility models that allow for the derivation of a cardinal utility function over some set of alternatives F_i after having obtained the stochastic preference relations between its members.[24] I use a model that imposes

[23]See Schuman and Presser (1996) for an introduction to these problems.

[24]One important model that I will not utilize is from Luce (1959). This model is more demanding than what I will use, but produces ratio-scale utility values instead of interval scale values. I think that the conditions it places on the structure of individual preferences is unrealistically demanding. Therefore, I will obtain ratio-scale significance by making assumptions about how measured preferences relate to general patterns in individuals' total set of preferences within some choice context. I will describe this in section 5.2.

relatively weak requirements on the structure of stochastic preferences between alternatives and yet still allows for the production of a cardinal utility function over them. Specifically, this need only produce an *interval*-scale utility function; in the literature these have been called *strong utility models*. Suppes et. al. (1989) present an axiomatization of such a model that derives differences in utility assignments from P-values. However, such a method only allows for meaningful distinctions in utility differences between *non-perfectly* stable alternative pairs. Therefore, Suppes et. al. also prove that local sets of non-perfectly stable alternatives can be strung together to obtain an interval-scale utility function over an entire set of alternatives (F_i), even when there are some perfectly stable preference-relations between some of those alternatives.

A scale is formed by defining a *difference structure* for F_i. This structure includes both an order of alternatives as well as an ordering of utility *differences* between those alternatives. In such a difference structure, some real valued utility function u over the set of alternatives under consideration $\{x, y, z, a, b, c\} \in F_i$ can be defined, and claims such as the following are possible:

$$u(x) - u(y) \geq u(y) - u(z)$$

Furthermore, these differences are determined by P-values, such that the following is true:

$$u(x) - u(y) \geq u(z) - u(a) \iff p(x \succ y) \geq p(z \succ a)$$

With all alternatives $\varphi \in F_i$ and $x \succeq y$ *iff* $xRy \wedge p(x \succ y) \geq \frac{1}{2}$ (where 'R' is some binary relation), a unique strong utility model over some set of alternatives F can be obtained *iff* the following conditions are satisfied:[25]

Comparability *If* $x \succeq y$ *and* $x \succeq z$ *then* yRz and *if* $y \succeq x$ *and* $z \succeq x$ *then* yRz.

Monotonicity *Suppose* xRy, yRz, xRz, aRb, bRc. *If* $p(x \succ y) \geq p(a \succ b)$ *and* $p(y \succ z) \geq p(b \succ c)$, *then* aRc *and* $p(x \succ z) \geq p(a \succ c)$. *Furthermore, if either antecedent is a strict preference relation, then so is the conclusion.*

Solvability *If* xRy, xRa, *and* $p(x \succ y) \geq t \geq p(x \succ a)$ *for some* $t \in (0, 1)$, *then there exists* $z \in F_i$ *such that* xRz *and* $p(x \succ z) = t$.

Connectedness *For any* $x, y \in F_i$ *that do not satisfy* $p(x \succ y) = \frac{1}{2}$, *there exists a sequence* $x = x_1, x_2, x_3 \ldots x_n = y \in F_i$ *such that exactly one of the following conditions holds* with $i = 0, \ldots, n-1$: (i) $x_i \succeq x_{i+1}$ *or* (ii) $x_{i+i} \succeq x_i$.

The *Monotonicity* and *Solvability* conditions produce *local* interval scales between those alternatives that do not form perfectly stable P-values. The *Comparability* condition ensures that middle alternatives (in the above case, y, z) are comparable with one another. The *Connectedness* condition ensures that all local interval scales can be extended and joined throughout the entire set of measured alternatives. This condition forces unique uni-dimensionality through the entire set F_i rather than just within each internally comparable subset of F_i.

Assume that the stochastic preferences are consistent with the *strong utility*

[25]This axiomatization comes from Suppes et. al. (1989), pg. 392.

model. According to this model it is possible to find an individual interval utility scale over all measured alternatives in F_i from the set of individual stochastic preferences over that set. This demonstrates that it is possible to obtain individual cardinal utility scales utilizing only sets of individual pair-wise judgments. As I argued above, such judgments are less problematic than estimates of individual interests or welfare, and so their use in the measurement of attitudes helps diminish worries about the subjectivity of individual utility measures. The *strong utility model* provides a set of conditions that must be satisfied by the method for finding individual utility scales, but it does not specify how exactly scale values should be calculated. Therefore, I will have to provide a scaling method that is compatible with the model. I will present one such method based on Thurstone's "Law of Comparative Judgment" (Thurstone, 1959, chapter 3). Although this is certainly not the only scaling method adequate to the task at hand, its mathematical details are relatively conceptually clear.

Thurstone's scaling method can be broken down into two main steps. First, response proportions (P-values) are transformed into standardized scale differences between all alternatives. Then all comparisons of alternatives are weighted based upon the expected level of instability from transient features (T) of each comparison. These weighted comparisons are finally averaged for each alternative to obtain utility scale values. In section 4.4.2, I present an example of such a scaling procedure, and therefore it may help to read that section concurrently with this one.

Referring back to figure 4.4 (pg. 123), in the mathematically simplest case, one can assume that the amount of instability created by different alternatives' Ts

are identical. This assumption eliminates the need for unique weights on each pairwise comparison, which is required in the more general case. I am making this assumption mostly for reasons of conceptual clarity, not because it is necessary to produce a scale.[26] I also assume that the effects of Ts are normally distributed, such the P-values produced by any utility difference can be found via the standard cumulative normal distribution function (Φ) as shown in equation 4.1.[27] This is Thurstone's *Law of Comparative Judgment* in its simplest form ("*Case 5*"), which determines measured P-values based upon underlying scale value differences between alternative pairs (φ_1, φ_2).[28] The normal form of the function Φ is determined by the normality of the term T.

$$p(\varphi_1 \succ \varphi_2) = \Phi[u(\varphi_1) - u(\varphi_2)] \tag{4.1}$$

However, in order to determine scale differences between two alternatives ($d(\varphi_1, \varphi_2)$) from some P-values, the inverse cumulative normal distribution (Φ^{-1}) must be used; this is also called a *probit* function in statistics.[29] The method for finding utility differences from P-values is shown in equation 4.2.

$$d(\varphi_1, \varphi_2) = u(\varphi_1) - u(\varphi_2) = \Phi^{-1}[p(\varphi_1 \succ \varphi_2)] \tag{4.2}$$

[26]Thurstone's "*Case 4*," for instance, offers a reasonably simple mathematical structure while allowing for differences in the Ts of each comparison.

[27]Just as with the standard normal distribution, this assumes a mean of 0 and a standard deviation of 1.

[28]For various relaxations of the assumptions required for the scaling method I use here see Falmagne (1985), chapter 5.

[29]The probit function (like all quantile functions of which it is a type) is difficult to use, and requires rather advanced mathematical methods to make calculations. Luckily, any standard spreadsheet or statistical package provides a function for carrying out such calculations, because they are quite common in statistics.

Equation 4.2 only finds utility *differences* between all alternatives, and does not directly produce a utility *scale*. Such a scale can be formed by calculating the average difference between each alternative and all other alternatives; if equality of all Ts is *not* assumed then this must be a *weighted* average. If the conditions of the *strong utility model* are satisfied, along with the additional assumptions I have just made, then this scaling procedure will produce *interval-scale* individual utility scales over all alternatives within the choice domain (F_i). This does not yield a zero point that is meaningful across individual utility scales or allow for immediate interpersonal comparisons of utility. However, it is not an arbitrary assignment of cardinal utility differences. The assumptions I have made about the normal distribution of dispersion effects of Ts in the above attitude model and the basic similarity of those effects across attitudes provide the non-arbitrary bases for making these *individual* utility assignments.

Importantly, the above method is able to produce cardinal utility scales making use of only patterns of pair-wise judgments (or choices) between alternatives. Thus, this does not require individuals to provide utility values *themselves*, but rather derives utility values from the choices of individuals. There is no entry for individual subjective impressions of individuals' own utilities, which are major sources of the difficulty in interpreting individual cardinal utility measures. The task of the next chapter will be to describe a non-arbitrary basis for making meaningful *interpersonal* comparisons of utility. Psychological characteristics shared by all individuals that allow for assumptions about the patterns of individual utility assignments will serve as this basis. However, in the next section, I will provide an example of how

individual cardinal utility scales can be used to represent individual attitudes over some set of policies.

4.4 An example from health care reform

4.4.1 Policy alternatives and attitudes

In this section, I will introduce an example that reveals how a cardinal utility model can be used to represent public opinion. I will continue this example in the next chapter (section 5.3). In this first part of the example, I will show how individual utility functions over some set of policy alternatives can be found from sets of stochastic preferences. Assume that there is some constituency (or society) $\mathcal{I}$ composed of a set of individuals $\{I_1, I_2, I_3\}$ who are attempting to decide between alternative public health care policies. Define a set of alternatives $\{x, y, z, a, b, c\} \in F_i$, where the alternatives have the following identities:

x Nationalized health care system

y Free health clinics for the poor

z Subsidies for full private health insurance for the poor

a Subsidies for catastrophic health insurance for the poor

b More tax deductions for health care costs

c Status quo[30]

[30]Assume that the society in question has a health care system similar to that in the United States prior to the 2009 reforms.

These identities will play no important role in the formal analysis of the individual and social preferences, but they should give the reader an understanding of the sorts of policy alternatives that might be modeled using the method I present here.

In the ideal case, each individual is asked a series of questions that measure the respondent's preference for every alternative relative to every other alternative. As I explain in section 4.1, the frequencies of responses are derived from a series of comparisons of each alternative pair. To prevent the tendency of individuals to respond consistently to identical question wordings, preferences about a single pair of alternatives might be measured through a series of differently worded questions comparing those alternatives. For instance, the pair x and z might be compared in two separate trials (t_1 and t_2), using the following questions:

$t_1(x, z)$ In order to improve health care access for the poor, would you rather the federal government completely nationalize the health care system or would you rather it give poor individuals subsidies to buy private insurance on the current market?

$t_2(x, z)$ In order to improve health care access for the poor, would you rather that the federal government give individuals who cannot afford private health insurance subsidies, so that they can purchase insurance plans on the current private market, or rather the federal government eliminate the entire current private health care system, giving all individuals equal access to all health care facilities regardless of their ability to pay?

In a complete measurement of attitudes, each individual would answer questions in many such trials for each alternative pair, with each trial differing in the precise manner in which individuals are asked about the pairs of alternatives. Each trial would introduce different transient considerations into the individual's response (elements of T), as I described in section 4.2.

Next I will show how a utility scale, satisfying the *strong utility model*, can be derived from some set of P-values.

4.4.2 Utility scaling

Using the standard cumulative normal distribution function from equation 4.2 (section 4.3), one can find utility differences between all policy alternatives from a set of P-values. I give a set of P-values for one individual in figure 4.5. Each cell of this table contains the P-values for $p(\varphi_1 \succ \varphi_2)$, where φ_1 is listed in the top row and φ_2 is listed on the leftmost column. Therefore, P-values for alternatives should be read down *columns* and not across rows. For purposes of demonstration, assume that this individual would rank the policy alternatives as follows: $x \succ y \succ z \succ a \succ b \succ c$. Throughout this example, I will name the individuals based upon their response patterns. I intend this to help the readers conceptualize the individual and social utility scales that will result from the scaling process. Given the definitions I have provided the alternatives, this individual might best be called "*The Socialist.*"

Utility *differences* are then calculated from P-values. For instance, consider the proportions obtained for z in figure 4.5. Substituting z for φ_1 and y for φ_2 in

equation 4.2, yields $\Phi^{-1}[p(z \succ y)] = \Phi^{-1}[0.4] = [u(z) - u(y)] = d(z, y) = -0.25$. Similarly, $d(z, x) = -0.84$ and $d(z, a) = 0.52$. These values, along with all other utility differences, are shown in figure 4.6.

φ_2 \ φ_1	x	y	z	a	b	c
x	—	0.25	0.2	0.1	0.001	0.001
y	0.75	—	0.4	0.25	0.05	0.001
z	0.8	0.6	—	0.3	0.1	0.05
a	0.9	0.75	0.7	—	0.25	0.2
b	0.999	0.95	0.9	0.75	—	0.3
c	0.999	0.999	0.95	0.8	0.7	—

Figure 4.5: P-values ($p(\varphi_1 \succ \varphi_2)$) for *The Socialist* ($I_1$)

φ_2 \ φ_1	x	y	z	a	b	c
x	—	-0.68	-0.84	-1.28	-3.09	-3.09
y	0.68	—	-0.25	-0.67	-1.64	-3.09
z	0.84	0.25	—	-0.52	-1.28	-1.64
a	1.28	0.67	0.52	—	-0.67	-0.84
b	3.09	1.64	1.28	0.67	—	-0.52
c	3.09	3.09	1.64	0.84	0.52	—
$u_1(\varphi_1)$	1.80	1.00	0.47	-0.19	-1.23	-1.84

Figure 4.6: Scale difference values ($\varphi_1 - \varphi_2$) for *The Socialist* (I_1)

For simplicity, I assume that no comparisons yield P-values of 0 or 1, and that the number of comparisons is very large. This allows me, in section 5.3, to present a case in which some individual exhibits very stable preferences for some alternatives as well as very unstable preferences for other alternatives. In actual measurement situations it will be nearly impossible to obtain so many comparisons of the alternatives from all individuals, and so it is likely that some individuals will

exhibit P-values of either 0 or 1. There are a number of ways to interpret such data consistent with the procedure I have given for obtaining scale values, but I will not deal with this in depth here.[31]

An averaging procedure over each column of utility difference values for each alternative in figure 4.6 produces the scale values (found in row $u_1(\varphi_1)$) for each alternative. This process produces *interval*-scale utility values, but utility *levels* and zero-points are arbitrary, and so individual utility values cannot be directly aggregated in a meaningful way to produce social utility values. In order to make the meaningful comparisons of utility necessary to aggregate individual utility values, some method must be used to transform the scaled values of all individuals to new scales that are interpersonally comparable. I will do this in the next chapter, and then continue this example in section 5.3.

[31] One obvious solution is to exclude all P-values of 0 or 1. This is Thurstone's (1959) own method in his classic studies of attitudes, in which he employs a scaling method similar to the one I use here. However, excluding these extreme values will compress the resultant utility scale. In light of this, it may be better to specify upper-bound P-values, with all P-values above that bound being assigned the upper-bound value. This makes it safe to claim that the utility difference observed for each alternative is *at least as large* as that which would obtain through some arbitrarily large number of observations. This is a significant choice that must be made by any empirical researcher, but luckily it will not change any conclusions of conceptual significance.

Chapter 5: Cardinal Public Opinion and Interpersonal Comparisons of Utility

In this chapter I will show how the *individual* utility model that I developed in the last chapter faces the well-known problem of making interpersonal comparisons of cardinal utility when those scales are aggregated. Such a social utility function is identified with *public opinion*, which I have argued should be thought of as having cardinal-scale measurability. The problem of how to find the opinion of a public that is relevant to the decisions of representatives then turns out to be a special case of the problem of making interpersonal comparisons of cardinal utility. The problem of how to make non-arbitrary (or "meaningful") interpersonal comparisons of utility strikes many people as being intractable, at least within many contexts. Therefore, many readers may at this point have skepticism of this entire project, at least insofar as it seems to rely upon the possibility of interpersonal comparisons of utility. However, as I will argue in this chapter, the direct consideration of that problem is necessary for an adequate explanation of how public opinion can be used by political representatives. The solution I will present is also instructive for other measurement domains in which interpersonal comparisons are necessary. I address this in section 6.2.3.

5.1 Interpersonal comparisons of utility

The most basic type of interpersonal comparisons of utility determines whether some individual is made better-off (or worse-off) to a greater degree by a social decision than another person. *This* can be determined using purely ordinal measures of utility, and in fact it is possible to obtain *full comparability* using a purely ordinal utility theory. However, such *full comparability* of *ordinal* scales requires that it be possible to produce a single ranking of all utility assignments for all individual utility scales such that for any two scales (u_1 and u_2) and any two alternatives (x and y), something like the following statement is meaningful:

$$u_1(x) > u_2(y) > u_1(y) > u_2(x)$$

It is unclear, however, how one might go about justifying such an ordering; it certainly could not be accomplished without substantive (and perhaps dubious) assumptions about the meaning of individual measurements. The use of *empathy* in order to infer such orders has received widespread attention among philosophers and economists.[1] This strategy largely relies upon the ability of individuals to infer the welfare of other individuals when they are in various states, typically by utilizing their own experiences of such states. It is plausible that a systematic empathetic process might yield very accurate predictions of the welfare produced by various social states, especially when the empathizer has a great deal of knowledge about

[1]This solution to interpersonal comparisons of cardinal utility has been given a large amount of attention in the literature. See Harsanyi (1955); Arrow (1977); Sen (1979); Goldman (1995).

the target of that empathy. As Amartya Sen (1970) argues, such judgments are commonplace. When I contemplate giving $100 to UNICEF it does not take great effort for me to estimate that the money will do a starving African child more good than it will do me. Such judgments become more difficult, but not different in kind, when welfare differences are not as great. Therefore, there is no great conceptual problem with interpersonal comparisons of utility. The problems that arise from such comparisons arise from problems in *measurement*, and there do not seem to be the same bases for *cardinal* interpersonal comparisons of utility that exist for ordinal ones. Finding such a justification will be the primary objective of this chapter.

My presentation of cardinal interpersonal comparisons of utility will require an understanding of various types of utility scales. I will primarily rely upon the classification system introduced by S.S. Stevens (1946), which has become the standard way of classifying measurement scales in the physical and social sciences. Scale types are typically defined by the types of transformations under which the scale remains invariant. For instance, ordinal scales remain invariant under *all positive* transformations, because no such transformations affect the *order* of utilities. However, for all cardinal utility functions the *differences* between utilities are significant metrics, and therefore only positive linear functions maintain invariance. For instance, the order of the utilities along a scale are unaffected by a positive exponential function, though that function changes the differences along that scale. A subset of cardinal scales, called *ratio* scales, are not invariant even under those linear scales that alter the zero-point. Cardinal scales that fail to be ratio scales are called *interval* scales.

Although the invararience conditions define scale types, those types also have

various significant metrics. Obviously, for ordinal scales only the orders of alternatives are of significance. For *interval* scales (and ratio scales), the ratios of utility *differences* are significant. To see this, consider two utility differences between x and y ($d(x, y)$) and y and z ($d(y, z)$). A meaningful metric for the interval scale containing all three alternatives is:

$$\frac{d(x, y)}{d(y, z)}$$

Any function that alters this value will also alter the utility scale; according to the definition of the interval scale, this value will not change as long as the transformation function is a positive linear one. It is typically said that such scales have significant utility *differences*, and while this is technically not true, because the *values* of differences are not significant but only the ratios of them are, I will use this language when discussing interval scales. Finally, notice that the significance of utility differences permits the establishment of a significant *unit* for that scale. Any utility difference between any two alternatives can be assigned the unit value '1', and the rest of the utility differences can be adjusted to maintain identical ratios of differences.

The *ratios* of utility *values* themselves are also significant for *ratio* scales. Consider the values of x, y, z; if these alternatives are along the same ratio scale, then statements of the following form are significant:

$$\frac{x}{y} : \frac{y}{z} : \frac{x}{z}$$

In fact, if these alternatives complete the scale, the above ratios totally characterize the utility structure of the scale. This requires that a *zero-point* has significance, something that is not true of interval scales.

The types of scales are related to the types of *interpersonal comparability*, though the type of scales do not *determine* the types of comparability between them. So, for instance, two different scales may both meet the requirements to be interval scales *internally*, but utility differences may not be significant *between* scales. For most purposes, it is enough that the *units* of cardinal scales be comparable. This allows for meaningful comparisons between utility *differences*, which makes the following sorts of statements possible:

Differences $u_1(x) - u_1(y) > u_2(y) - u_2(x)$

However, because these scales do not share meaningful zero-points, it is *not* meaningful to make the following sorts of statements:

Order $u_1(x) > u_2(y) > u_1(y) > u_2(x)$

Product $u_1(x) > u_2(y) * 2$

Two interval scales, therefore, may share units but yet it may be impossible to order utility values of those scales along a single scale, even though such an ordering is trivial to form *within* each scale.

A stronger form of comparability, called the comparability of utility *levels*, allows for meaningful statements about utility values in addition to differences. This makes significant all three types of statements from above (Difference, Order, and

Product). Comparability of utility levels allows for the placement of the utility values of individuals along a single linear scale, because there is some meaningful point along which all scales can be aligned. Comparability of *units* does not allow this, because it only gives information about the utility values relative to other utility values within each *individual* scale. Unit comparability allows one to make judgments of overall utility differences between two alternatives across some set of comparable individual scales, but it does not allow for any operations beyond addition (and subtraction). For instance, it is not possible to say that the utility some individual I_1 assigns to x is twice as great as what some other individual I_2 assigns to y, because there is no meaningful zero point shared by their two utility scales. This precludes the use of an averaging procedure to aggregate utility values. Some reflection will reveal that it is actually quite difficult to obtain unit comparability without substantial knowledge, or at least assumptions that presuppose some knowledge, about the systems being measured. In fact, this sort of comparability does not exist in some rather mundane examples from the natural sciences.

The problem of how one makes meaningful comparisons between various measured scales arises in nearly every domain of science. Typically, scientists overcome this problem by making basic assumptions about all individual data *sources*; such assumptions allow for the statistical inferences that characterize most scientific findings. In the case of psychology, certain assumptions are usually made about the cognitive structures of *all* individuals. These are assumptions *specific* to particular theoretical contexts, and possibly not applicable in others. A simple case will illustrate this. Consider two thermometers that both measure temperature in Celsius,

but neither of which are perfectly calibrated. In other words, the output of the thermometers are simple *translations* of one another, so that the output of one is equal to the output of the other plus some constant value. Given these assumptions, it can be assured that comparisons of the magnitudes of *differences* on these scales are meaningful; the *units* of the thermometers are the same. However, comparisons of scale *levels* are not meaningful, because the thermometers have different zero-points.[2] In other words, the zero-points of the two thermometers refer to different levels of kinetic energy when used to measure the same system. However, comparisons of scale differences between either of these thermometers and a third thermometer that measures temperature in Fahrenheit, or some other unit, is not meaningful without the use of a function that *transforms* either the Fahrenheit or Celsius scales. Such a function applies both a constant translation to the scale as well as some multiplier to account for differences in the units.

The proper transformation required to compare two thermometers is easy to apply, as long as it is assumed that they are indeed *working* thermometers. Given two properly working thermometers (u_1 and u_2) then, the measurement of any two states (x and y) by those thermometers can form the basis of a unit transformation between those scales. However, this only arises because the physical theory about temperature and working thermometers allows for a simple transformation function between scales, when some unit multiplier m is known, such that:

[2]See Bossert and Weymark (2004) for a overview of the various forms of utility scales and their respective comparability.

$$\frac{u_1(x) - u_1(y)}{m} = u_2(x) - u_2(y)$$

Additionally, m can be found from the measurement of any two states x and y through simple algebra, because it can be stipulated that for any state y, $u_1(y) = u_2(y) = 0$ (where 0 can also be any other value). Of course, nothing like this simple function exists in the case of any standard utility theory. There is no expectation that some individual's valuation of an alternative must be related to another individual's valuation of that same alternative in the same way that it is expected that two working thermometers agree on the measurement of some single energy state. Therefore, it is unlikely that comparisons of utility scales will be so straightforward. However, less straightforward scale transformations common in other domains offer a clue as to how meaningful comparisons of individual *utility* scales might be made.

Another example will show how *context-specific* assumptions are used within the social sciences to accommodate the often large differences in baseline individual response tendencies in psychological experiments. Consider a case in which an experimenter wants to determine whether some training technique improves reading speed. The simplest method for measuring this is to randomly divide a sample of subjects into two groups, those who receive the extra reading training (the "Manipulation" condition) and those who only receive an unrelated sort of training (the "Control" condition). All subjects can then be asked to read some passage with the expectation that they will later be tested for reading comprehension, and the

number of words that they read in some interval of time can be measured. Assume that the experimenter obtains data from 12 subjects (one for each cell), which is shown in figure 5.1.

Column	1	2	3	4	5	6	Mean
Control	72	62	53	100	91	46	70.7
Manipulation	82	65	60	105	92	50	75.5

Figure 5.1: Number of words read

One standard method to determine whether some manipulation has an effect on performance is to use a statistical hypothesis test called the "*t*-test." If the sample means of two conditions are different enough (and/or error small enough), then the *t*-test *rejects* the statistical hypothesis that the means of their *populations* are identical. Though, given that this is a *statistical* test, there is still some small probability (often .05 or .01) that the test rejects the hypothesis when in fact the population means are identical. In the example from figure 5.1, even though the mean value for the *Manipulation* is greater than that of the *Control*, a *t*-test finds there to be no statistically significant difference between the two conditions; the variability between *individual* reading speeds, regardless of the condition, overwhelms any systematic difference between the *conditions*.[3]

The problem of large variability between individuals, unrelated to differences in the two conditions, can be avoided by testing each individual in *both* conditions. Sometimes this is not appropriate, but by repeating the two conditions for all subjects the systematic variability between *subjects* is eliminated from the analysis of

[3]The probability that the *Manipulation* condition does *not* produce higher scores than the *Control* condition is 0.69 with a two-tailed *t*-test, assuming equal variance in the two conditions.

systematic differences between the *conditions*. In the above case, treating each column in figure 5.1 as a *subject* will represent such a repeated analysis for six (instead of twelve) subjects. A t-test of that experiment shows that subjects read significantly more words in the *Manipulation* condition than in the *Control* condition.[4] The key for my purposes is that a repeated experimental structure allows for the interpretation of the experimental results (the *Manipulation* condition) within the context of each subject's responses from the *Control* condition. This is a very simple example of using individual response tendencies to *interpret* other responses by that individual, and even with only two data points such a procedure permits a much more powerful analysis of the experimental results. However, the results of these statistical methods are only meaningful once some basic assumptions about the source of the data are made; without such assumptions it is largely impossible to make inferences in most scientific contexts. For instance, the t-test requires that random error be distributed normally and for measurements to be independent. This general theme, that assumptions plausible within a specific theoretical context allow for a better interpretation of data, will guide the construction of the public opinion model I provide here.

A related type of data interpretation is the *normalization* procedure. One simple form of normalization that many are familiar with is the "curving" of exam scores. Consider a case in which one wants to know whether a student did "better" (represented with '$\succ$') on the midterm than on the final exam. Assume that both

[4]The probability that the *Manipulation* condition does *not* produce higher average scores than the *Control* condition is 0.01 with a two-tailed t-test, assuming that the two conditions are repeated.

exam scores are out of 100 possible points: the score of the midterm $(m) = 70$ and the score for the final $(f) = 55$. The simplest method for interpreting "better" in this context is to use the "raw" score, so that in this case $m \succ f$. However, as anyone who has ever given or written an exam can attest, this is also probably the wrong interpretation. Rather, knowledge about the relative difficulties of the exams should be used in interpreting the scores, and usually the best method of obtaining this information is by inferring it from the test scores of *all* tested students. One common method of interpreting exam scores is to assign them *grades* based upon score percentile; although a normalization procedure could also assign normalized scores to raw scores, it will be simpler to examine how normalized grades can be assigned to raw scores. For instance, if the midterm has a mean score of 71 and the final has a mean score of 40, then under most standard methods for assigning *grades*, it is likely right to say that $f \succ m$. In the standard case of grade "curving," the letter grade assigned to an exam score is determined by the percentile of the *population* of scores that the exam score lies above. For instance, an 'A' may be assigned to scores in the top 13% of the population of exam scores, a 'B' to the top 40% of scores that are below the top 13%, and so on. Importantly, the statistical features of the *population* of test scores is not directly accessible, but can only be inferred from the sample of scores, where the sample is the set of students who wrote the exam in any given case. More precisely, it must be assumed that the students who take some exam after attending some class are a random sample from the population of possible students. There should be nothing unique about the set of students who actually attended the class and wrote the exam; they are no better

or worse students than some other possible set of students. If this condition is not satisfied, if say there is reason to think that the set of students who wrote the exam in this case is worse than the population of students, then there would seem to be no reason to normalize scores using the "curving" procedure such that scores improve; it would be correct to interpret those students' scores as worse than average, and so it would be correct to assign them lower than average test scores. The structure of this example is important, and I will use a related method to make comparisons between utility scales of public policy attitudes. However, one should also draw a more general conclusion from this example: normalization techniques are widely used to allow for meaningful comparisons between measurement scales, even though such processes (such as exam score "curving") are rarely thought of as methods of making inter-scale (or interpersonal) comparisons.

My examples up until this point have dealt with scales that are typically not measured using "utility," but some may argue that utility theory presents a special problem for the comparison of scales. Interpersonal comparisons of utility in welfare economics and social choice theory have typically been treated interchangeably. There is some good reason for this, because the purely formal findings relevant to interpersonal comparisons are applicable to both fields. However, the theoretical basis for utility in welfare economics and social choice theory are importantly different, such that it may be possible to make sense of interpersonal comparisons of utility in the context of social choices even when it may not be possible in the context of welfare economics. In the case of individual *welfare*, the sort of assumptions that would be required for interpersonal comparisons of utility are certainly complex. As

has often been argued, there seems to be little grounding for measurements that would yield comparable scales, because personal welfare is too complicated and inherently subjective.[5] In contrast to welfare economics, in social choice theory the term 'utility' need only be considered a term of art; it refers to any cardinalization of individual preference sets, but generally does not bring with it any specific interpretation.[6] In this *social choice* context, claims about utility differences represent differences in the magnitude of *preferences*. They need not imply that some choice would actually improve the welfare (or any other notion of utility) of the individual, and it is likely implausible to say that welfare is fully determined by preference satisfaction.

'Preference' itself is a technical term in social choice theory, and normally its precise interpretation can be left out of any social choice analysis. However, a relatively specific interpretation of this concept is necessary for one to make substantive assumptions about the *source* of utility values, and the method I present here will depend upon certain assumptions about this source. As should be obvious from the previous chapters, I will interpret preferences as representations of individual *attitudes* about alternatives. Individual attitudes, when characterized in a particular way, are simple enough entities that they allow for straightforward *interpretation* and eventual comparability. This distinguishes them from individual welfare or interests, which are typically characterized as having complex structures, the specifics of which are of great controversy. There is comparatively a great deal of agreement

[5]Such a view has been frequently expressed, but see Robbins' (1952) famous presentation of it.
[6]For an overview of the use of 'utility' in social choice theory and economics see Broome (1999).

about the basic form of attitudes, and especially the attitudes about public policies (see my discussion in 4.2).

As in the case of temperature scales and exam grades, meaningful comparisons between different utility scales require some understanding of the underlying phenomenon that produces those scales, or at least the determination of some relationship between them. Arrovian social choice theory, which excludes interpersonal comparisons of utility, is premised on there being no *necessary* relationship between individual utility scales, but this is only one possible relationship. I will argue that the public policy attitudes of most individuals are related to each other in some describable way.

5.2 Interpersonal comparisons and social utility

In this section, I will impose a number of assumptions on the *source* of individual utility assignments to permit similar interpretations of all individual utility values within a *choice context* (D_j).[7] In terms of attitude strength, I impose assumptions about the underlying sources of individual attitudes in order to provide similar interpretations for measurements of individual attitude strengths. When this interpretive framework is applied across individuals, this allows for ratio-scale measureability of individual utility assignments and full ratio-scale interpersonal comparisons of utility between individuals.[8] Although these assumptions are given

[7]This is identical to D_i, but in this chapter I will use the index 'i' for u_i

[8]Note that this ratio-scale measurability arises from individual utility data that only has interval-scale measurability. The underlying assumptions then will provide both additional structure to individual scales as well as provide a means to compare them.

in terms of utilities themselves, they will be justified by making reference to psychological processes, and specifically the structure of attitudes. I will begin with a number of simplifying assumptions about the basic structure of the space of utility assignments. Assume that each individual's set of utility assignments over *all possible* alternatives (D_j) within some context of choice can be described by some frequency distribution of alternatives in D_j over possible utility assignments.[9] Figure 5.2 shows two such distributions, with the x-axis equal to possible utility assignments and the y-axis equal to the frequency of alternatives being assigned each utility value. These distributions can be thought of as *dispositions* to assign various utilities. Next, assume that the set of alternatives under consideration (F_j) is a *Random Sample* from D_j, such that the probability of a randomly selected alternative having any given utility can be given by some probability distribution over the set of utility assignments. This is a standard set of assumptions needed for some descriptive statistics of a *sample* to approximate those of a *population*. I take the *Random Sample* condition to be uncontroversial. It does not mean that the alternatives must be chosen in some random fashion, but just that the utility *values* an individual assigns to them are not used in any non-random selection process of the alternatives under consideration.

My goal will be to show how full ratio-scale interpersonal comparisons might be made meaningful. Up until this point, I have argued that *individual* utility differences are meaningful, because they are derived from a behavioral method. I have

[9] A choice context contains a set of alternatives that are comparable. These contexts will likely be rather large, containing perhaps all public health care policies or tax policies, but will exclude many alternatives. For instance, it may not make sense to compare health care policies to foreign policies, and so they may be in different choice contexts.

also claimed that the attitudes of all individuals are structured in roughly similar ways, consistent with figure 4.4 (pg. 123). This assumption might be tightened by a claim that individuals have *identical* attitude structures, though this is something I will eventually reject. According to the tightened assumption about the structure of individual attitudes, all that differs between individuals are the actual utilities they assign the alternatives; their Ts (from figure 4.4) and their general utility assignment dispositions are identical. Nonetheless, zero-points of individual utility scales would still be arbitrary, because these are the result of arbitrary scaling decisions and did not arise from individual behaviors. By assuming that the population of alternatives D_j forms a *uniform distribution* over utility assignments, and the set of alternatives under consideration F_j is sampled randomly from D_j, then the *mean* of $u_i(F_j)$ approximates the mean utility value of the population $u_i(D_j)$. All individual utility scales can then be *translated* such that the mean utility value for each individual scale is aligned. After such a translation, a simple *averaging* procedure will be fully sensitive to the utility information available within the individual utility functions. Call this the *translate and average* procedure.

The use of *translate and average* is based upon indefensible structural assumptions, because there are likely to be individual differences in the ability of individuals to express their underlying utility differences in pair-wise choices. In terms of the attitude model shown in figure 4.4, it is likely that Ts will differ between individuals, or at least there seems to be little reason to think that they will not vary. But it turns out that an equally difficult problem exists with the sort of *normative* assumption that must be made for *translate and average* to be appropriately involved in the

social decision process. Call this assumption *Expert Preference*: *"experts"* (those with smaller Ts, and thus less error per choice situation) and those with *overall stronger attitudes* should both have more overall say in the social decision. The case in which all Ts are identical across individuals allows for a special case of *Expert Preference*, in which there are no "experts" and only those with stronger attitudes, but the core problem remains. Both versions of *Expert Preference* violate norms of basic fairness that are central to democratic theory. More precisely, *translate and average* seems to violate a principle that all individuals must have equal *overall* influence on social decisions.[10] Later in this section, I will show that this principle can be captured by a normative assumption called *Overall Fairness*.

Consider a new method called *normalize and average*. It retains *Random Sample* from the *translate and average* model, but it assumes that each individual population of utility values is distributed *normally* (instead of uniformly) and that both the mean and variance of those normal distributions must be used to properly interpret individual utility assignments. According to *translate and average* only the mean of $u_i(D_j)$ is a significant statistic, but under *normalize and average* the variance is also significant. Additionally, the utility assignments over F_j (the sample) can be used to determine both the mean and variance of the distribution. This has the result of significantly constraining the meaningfulness of individual utility differences from the scaling method alone; in order to determine the interpersonally meaningful differences, one must make a further assumption about the *variance* of

[10] A similar principle also holds for Range Voting, which allows individuals to give candidates cardinal scores that are then aggregated to determine a social decision function. See Pivato (2013).

the population of all utility assignments by that individual.

The *normality* of individual populations of utility assignments is justified by assuming that each individual is *disposed* to make certain types of utility assignments. This psychological fact about each individual permits a proper interpretation of individual utility assignments. In the psychological literature about attitudes these are called "dispositional attitudes" (Hepler and Albarracin, 2013). Research involving these dispositional attitudes typically classifies individuals into categories based upon their dispositions to form certain types of attitudes. For instance, the tendencies of individuals to form negatively valenced or positively valenced attitudes might differ between individuals. This can be easily represented in terms that I have already introduced. Assume that the dispositions of individuals can be characterized by normal probability distributions, with specific means and standard deviations. Distributions may also differ in their skewness, such that a normal distribution with a positive skew, like that shown in figure 5.3, represents an individual disposed to having more negatively valenced attitudes; it might be best to call this individual a "pessimist." In terms of utility, the individual is disposed to assign "below average" utilities. The inclusion of skewed-normal distributions in the set of possible attitude representations increases the generality of the attitude model. However, for simplicity, I will assume that all individuals are neutrally disposed, such that their dispositions can be represented by non-skewed normal distributions, which can differ in their means and variances. In other words, individuals differ in the size of the utility differences they tend to assign.

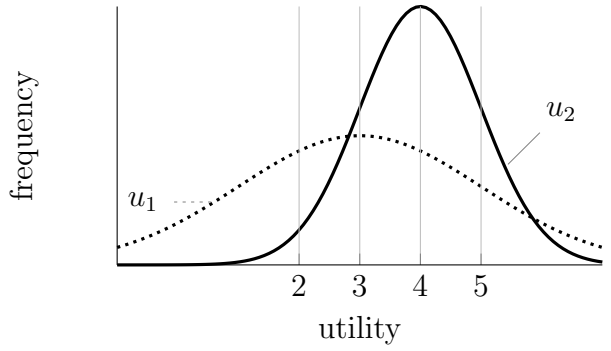

Figure 5.2: Neutral dispositions

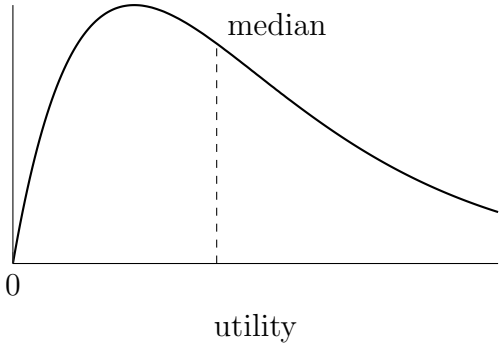

Figure 5.3: A "pessimist"

Given the fact that F_j (the alternatives under consideration) is a random sample from D_j, the mean and standard deviation of the utility assignments over F_j can be used to estimate those of the normal distribution of a individual's disposition. For instance, consider two individual dispositions for utility assignments (u_1 and u_2) over D_j shown in figure 5.2. Four utility *values* $(2, 3, 4, 5)$ are plotted on the x-axis.[11] The differing dispositions against which u_1 and u_2 are interpreted means that utility differences in u_1 are less *significant* than those same utility differences in u_2. This is because the individual who exhibits u_1 tends to produce larger utility differences than the individual who exhibits u_2. Therefore, the difference between say 4 and 5 is more significant for u_2 than it is for u_1. This degree of significance can be represented as *normalization* transformations of u_1 and u_2; each utility function is replaced by a normalized function $t_i(F_j)$ for each individual i. These new transformed utility scales have meaningful zero points, which are the estimated means of the frequency distributions over $u_i(D_j)$. They also permit comparability of *units*, which turn out to be the number of standard deviations from the mean. Taken together, these

[11]Note that each value might be assigned to different alternatives for each individual, such that for example $u_1(y) = 3$ and $u_2(x) = 3$.

also bring about interpersonal comparability of utility values themselves, or full ratio-scale comparability. This renders meaningful the following sorts of statements about the utility values in figure 5.2 where, for example, the transformed value of assignment 2 by u_1 is represented as $t_1(2)$:

1. $t_1(3) = t_2(4) = 0$
2. $2 * (t_1(5) - t_1(4)) = t_2(5) - t_2(4)$
3. $t_1(5) = t_2(5)$
4. $-(t_2(5)) = t_2(3)$

The above mathematical statements are instructive about the level of comparison permitted by *normalize and average*. Zero points are placed at the inferred means of the populations of utilities (item 1), and then units can be compared based upon the standard deviation of each distribution. Item 2 shows that each difference in utility of 1 for u_2 is equivalent to 2 units of utility for u_1. With the comparison of zero points and units, this allows for comparisons of utility values (item 3). Additionally, meaningful zero points provide a meaningfulness to positive and negative values (item 4). This is full ratio-scale comparability. Notice that the structure of the transformed utility function ($t_i(F_j)$) is a result of the form of the distribution representing the dispositional attitude. For instance, a positively skewed distribution of an individual with a negatively valenced dispositional attitude (a "pessimist" from figure 5.3), would result in a transformation function that produces larger differences for equal changes to the right of the median than to the left.[12] Because the

[12]Skewed distributions introduce additional complexities in identifying the point of central ten-

individual is already more disposed to assign "worse than average" utilities to alternatives, utility differences between alternatives that are rated "better than average" are more significant than those that are rated "worse than average."

If the normalization method provides the basis for meaningful full ratio-scale interpersonal comparisons of utility, then an average of the normalized utilities could plausibly provide a significant measurement of the aggregate attitudes (the public opinion). This then provides a basis for the use of attitude strengths in the measurement of public opinion. In section 5.3, I provide an example of a normalization function that is able to represent the relative significance of utility differences and values in this way. It may help the reader to go through that example in order to understand what the normalization function might look like.

The structural assumptions on utility that I have just given are largely empirical, and so no wholly conceptual argument could make them any more than plausible; I certainly cannot *prove* that they hold in normal cases of public policy attitudes. Although there is support for the existence of certain dispositional attitude types in the empirical literature (Hepler and Albarracin, 2013), the existence of dispositional attitudes characterized by normally distributed utility assignments can also be supported by the following specific intuition: in most choice contexts people assess a few alternatives as being excellent, a few as terrible, with most alternatives falling somewhere in between. Such generalizations about restaurants, books, and wines (just to name a few) seem reasonable, but I think that this intu-

dency, because the mean and median (two standard measurements of central tendency) are not identical, unlike in the case with the non-skewed normal distribution. For a number of reasons, the *median* is likely a more appropriate measure of central tendency than the mean for such distributions, but I will not discuss this here.

ition extends to public policy alternatives as well. And note that in a more general theory it need only be possible to describe each individual as having *some* dispositional attitude, rather than just the neutral disposition that is represented by the normal distribution. In principle, any individuals with dispositions capable of being represented by some variant of the normal distribution (including skewed-normal) can be meaningfully compared to one another, with each individual's distribution serving as the basis for interpersonal comparisons. This consideration is in need of further elaboration, but I have only meant to show its general plausibility here.

The argument for the use of information about dispositional attitudes in the interpretation of utility scales can then be characterized as follows: in order for the *significance* of an individual's responses to be understood, something about her general tendencies (or dispositions) must first be known. For example, when someone who generally responds negatively toward alternatives — meaning that she has a negatively valenced *dispositional attitude* — reacts positively toward an alternative under consideration, this should be interpreted as being more significant than when someone who has a *positively* valenced dispositional attitude reacts in the same way. Similarly, when someone who usually does not exhibit strong attitudes does so, this is more significant than when someone who *normally* exhibits strong attitudes exhibits those same attitudes. There is a final point that gets to the core of the strategy I suggest here. The individual utility scales I produced in section 4.4 were *interval* scales, meaning that only the utility differences were meaningful; this is what is implied by the *strong utility model*, which serves as the basis of the scaling method. However, this means that the zero point is not given any special

significance, and so the scale does not permit meaningful multiplication or division, nor does it permit meaningful *full interpersonal* comparisons of ratio-scale utilities. One intention of the structural assumptions I have just given is to make the zero point of each utility function significant. In other words, these assumptions make significant the fact that some utilities are above and others below the mean utility value. This role for the utility assumptions is important, because it allows for a meaningful normalization procedure for the individual utility scales, even though such a procedure would not produce meaningful results if the utility values were *only* interpreted by way of the *strong utility model*. This is necessary because the specific normalization procedure I present in section 5.3 requires that individual utility values be divided by a measure of variance in order to obtain the normalized utility value, and the result of such *division* is not directly meaningful for interval-scale values.

The fact that something is interpreted as significant does not necessarily make it proper or just to use it in making a social decision. Such decisions must be guided by some normative (or moral) principles. And just as there is a *normative* assumption that coincides with the structural assumptions of *translate and average*, there are normative assumptions that coincide with *normalize and average*. Again these normative assumptions play two possible purposes. First, they show how structurally significant utility values are appropriate for use in social decisions. Secondly, they show how ambiguities in structural significance might sometimes be ignored on normative grounds.

The use of *normalize and average* can be supported by two normative assump-

tions about democratic social decision rules: *Special Interests* and *Overall Fairness*. *Special Interests* states that any individual who has a *stronger* attitude for some alternative x over some other alternative y should have more influence on the resultant social utility difference between those two alternatives than some other individual who has a weaker attitude toward those two alternatives. *Overall Fairness* requires that all individuals must have equal influence on the resultant social utility scale. The desirability of *Special Interests* requires the understanding of the source of utility differences within individual utility functions. The *strong utility model* that served as the basis for the utility scaling procedure I presented in section 4.3 derives utility differences from differences in measured preference stabilities (P-values). However, this in itself does not provide an interpretation of preference stability necessary to determine the relevance of such differences to public policy decisions. Luckily, the attitude theory presented in section 4.2 offers just such an explanation for this. If it is assumed that individuals are generally able to exhibit their underlying attitudes in similar ways between different pairs of alternatives (i.e. that Ts are similar for all comparisons by each individual), then stability differences represent significant differences in utility *within* each individual's utility function. But this may not permit the use of measures of *overall* differences in preference stability *between individuals*, unless it is assumed that Ts are the same for all individuals. I have already argued that it is implausible that the Ts of all individuals are identical, because this would deny the existence of individual differences in response tendencies. Because *Special Interests* does not rely upon this assumption, it does not fail in the same way as *Expert Preference* does.

The dynamics of the *normalize and average* method illuminates certain features of *Overall Fairness* that are important to its suitability in democratic decision making. According to *normalize and average*, as the utility an individual assigns to any alternative moves further from the mean of the set of the individual's utility assignments, its influence on the social choice function approaches a limit and the influence of the individual's other attitudes approaches zero. Although some individual or group may have more influence over the social utilities of some subset of the alternatives, this will always decrease influence of that individual or group over the social utilities of the other alternatives.[13] Additionally, for social decisions within large groups, the effect of *Special Interests* will be small when the sub-group possessing the strong attitude is small. Even when some large minority group exhibits very strong attitudes in favor of (or against) some alternative, the normalization procedure only gives that preference more weight as weight is taken away from the group's other preferences. It may be best then to think of a group or individual with strong attitudes as having weight *redistributed* from weaker attitudes to those stronger ones, rather than *more weight* being given to the stronger attitudes.[14]

In this section, I have not attempted to come to any firm conclusions about which principles should be chosen, but rather have intended to show how deliberation on that issue might go. In the next section, I will continue the example I began in

[13]The measurement of "influence" is complex and I will not deal with it in detail here. This can perhaps, crudely, be measured by the total absolute values of the differences contained within an individual utility function.

[14]This is similar in certain respects to the Adjusted Winner fair-division procedure (Brams and Taylor, 1996). However, whereas the Adjusted Winner procedure makes use of individual (subjective) *reports* of the relative importance of the various goods to determine the best distribution of goods, my model *measures* and *infers* the relative significance of individual preferences from choice behaviors (or judgments).

section 4.4, demonstrating how *translate and average* and *normalize and average* can be used to make cardinal measurements of public opinion. More importantly, I will show how they differ. Then, in section 5.4 I will give an argument for why *normalize and average* might be thought to instantiate the *general will* I introduced in chapter 2, thus making it an appropriate model for the *Duty of Political Responsiveness* I introduced in chapter 3.

5.3 An example of social utility from health care reform

A continuation of the example from the previous chapter will reveal how a normalization technique can produce interpersonally comparable utility functions that suitably represent public opinion, even when those individual utility scales arise from stochastic preferences of differing stabilities. Consider a utility function $u_i(F_1)$ for each individual $I_i \in \mathcal{I}$, from which a set of individual *transformation* functions $(t_i \circ u_i)(F_1)$ can be found. For simplicity, assume that all members of $\mathcal{I}$ are *Moderates*.[15] These transformation functions transform utility values for all individuals into normalized versions that take into account the mean and variance of each individual's utility values. After such transformations are complete for all individual utility functions, a social choice function can find meaningful cardinal social utility for all alternatives. For clarity of presentation, I will use an averaging procedure to find these social utilities, though any number of functions may be used. An averaging procedure has been preferred by many theorists, especially utilitarians,

[15]I described this condition and the more general conditions it depends upon in section 5.2.

for a number of reasons [16] Perhaps the most obvious is the equal consideration it seems to give all individuals, at least in its unweighted form. It seems to be a natural and simple starting point for any cardinal social choice function, and so I use it exclusively in my example here. However, my conceptual conclusions do not depend upon its use.

Figure 5.9 shows the set of utility functions for the individuals in $\mathcal{I}$, $U = \{u_1, u_2, u_3\}$ and the *normalized* versions of those functions $U_t = \{t_1, t_2, t_3\}$, all over the set of alternatives $F_1 = \{x, y, z, a, b, c\}$. Note that the individual utility scale u_1 is for the "The Socialist" that I introduced in section 4.4.2, figures 4.5 and 4.6, and I repeat the utility values for "The Socialist" in figure 5.4. The P-values and utility values for the individuals I_2 and I_3 are given in figures 5.5 to 5.8. Given their utility scales, it might be useful to call I_2 "The Risk-averse Centrist" and I_3 "The Apathetic Libertarian," in a similar sense as I have called I_1 "The Socialist."

φ_2 \ φ_1	x	y	z	a	b	c
x	—	-0.68	-0.84	-1.28	-3.09	-3.09
y	0.68	—	-0.25	-0.67	-1.64	-3.09
z	0.84	0.25	—	-0.52	-1.28	-1.64
a	1.28	0.67	0.52	—	-0.67	-0.84
b	3.09	1.64	1.28	0.67	—	-0.52
c	3.09	3.09	1.64	0.84	0.52	—
$u_1(\varphi_1)$	1.80	1.00	0.47	-0.19	-1.23	-1.84

Figure 5.4: Scale difference values ($\varphi_1 - \varphi_2$) for *The Socialist* (I_1)

"The Risk-averse Centrist" (I_2) is so named because she generally has "cen-

[16] For instance, see Harsanyi's (1955) utilitarianism.

φ_2 \ φ_1	x	y	z	a	b	c
x	—	0.9999999	0.999999999	0.99999	0.999	0.99999999
y	0.0000001	—	0.6	0.25	0.2	0.55
z	0.000000001	0.4	—	0.3	0.1	0.45
a	0.00001	0.75	0.7	—	0.25	0.75
b	0.001	0.8	0.9	0.75	—	0.85
c	0.00000001	0.45	0.55	0.25	0.15	—

Figure 5.5: P-values ($p(\varphi_1 \succ \varphi_2)$) for *The Risk-averse Centrist* (I_2)

φ_2 \ φ_1	x	y	z	a	b	c
x	—	5.20	6.00	4.26	3.09	5.61
y	-5.20	—	0.25	-0.67	-0.84	0.13
z	-6.00	-0.25	—	-0.52	-1.28	-0.13
a	-4.26	0.67	0.52	—	-0.67	0.67
b	-3.09	0.84	1.28	0.67	—	1.04
c	-5.61	-0.13	0.13	-0.67	-1.04	—
$u_2(\varphi_1)$	-4.83	1.27	1.64	0.61	-0.15	1.46

Figure 5.6: Scale difference values ($\varphi_1 - \varphi_2$) for *The Risk-averse Centrist* (I_2)

trist" policy attitudes, but is especially averse to radical changes like x (a nationalized health care system). Likewise, I_3 is named "*The Apathetic Libertarian*," because although she slightly favors libertarian-friendly policies (tax deductions and subsidies rather than nationalization and free services) she only prefers those alternatives a small degree more than other alternatives. Her stochastic preferences and utility values are shown in figure 5.7 and 5.8.

The utility scales can be transformed into *normalized* versions through a relativization of each utility scale to its mean and standard deviation. Assuming that all individuals are *Moderates*, the mean utility value of the scale becomes its zero-

φ_2 \ φ_1	x	y	z	a	b	c
x	—	0.4	0.45	0.6	0.65	0.55
y	0.6	—	0.55	0.7	0.75	0.65
z	0.55	0.45	—	0.65	0.7	0.6
a	0.4	0.3	0.35	—	0.55	0.45
b	0.35	0.25	0.3	0.45	—	0.4
c	0.45	0.35	0.4	0.55	0.6	—

Figure 5.7: P-values ($p(\varphi_1 \succ \varphi_2)$) for *The Apathetic Libertarian* (I_3)

φ_2 \ φ_1	x	y	z	a	b	c
x	—	-0.25	-0.13	0.25	0.39	0.13
y	0.25	—	0.13	0.52	0.67	0.39
z	0.13	-0.13	—	0.39	0.52	0.25
a	-0.25	-0.52	-0.39	—	0.13	-0.13
b	-0.39	-0.67	-0.52	-0.13	—	-0.25
c	-0.13	-0.39	-0.25	0.13	0.25	—
$u_3(\varphi_1)$	-0.08	-0.39	-0.23	0.23	0.39	0.08

Figure 5.8: Scale difference values ($\varphi_1 - \varphi_2$) for *The Apathetic Libertarian* (I_3)

point and one unit of normalized utility is equal to one of its standard deviations. Or formally, where F_i is the set of utility assignments over F by I_i, $t_i(\varphi)$ is the normalized utility assignment by I_i to alternative $\varphi \in F$:

$$t_i(\varphi) = \frac{u_i(\varphi) - \bar{F}_i}{SD(F_i)} \tag{5.1}$$

Equation 5.1 calculates what is called a *t-statistic*. Consider the normalized utility assignment by individual I_1 to alternative x ($t_1(x)$). The calculation of the normalized value for this alternative requires that the standard deviation ($SD(F_1)$) and mean ($\bar{F}_1$) are calculated for that individual's set of utility values. This yields the following values:

$$SD(F_1) = SD\{1.80, 1.00, 0.47, -0.19, -1.23, -1.84\} = 1.37$$

$$\bar{F_1} = \frac{1.80 + 1.00 + 0.47 + -0.19 + -1.23 + -1.84}{6} = 0.01$$

These values, along with the original utility assignment $u_1(x)$ are included in the normalization function to find the new, normalized utility assignment:

$$t_1(x) = \frac{1.80 - 0.01}{1.37} = 1.31$$

This process is repeated for all utility values of all individuals, which are shown in figure 5.9. I give the results of an averaging social choice function over the normalized individual utility functions in figure 5.10, column $avg(U_t)$. When using this normalization procedure, no individual is able to have greater *overall* influence on the resultant social choice function than any other. However, this social choice function is sensitive to larger differences in individual utility scale differences, which are the result of more *stable* stochastic preferences.[17] Therefore, although an individual can have more influence on the social utility assignment of any specific alternative(s), this will in turn increase the variance of her utility assignment set and decrease the utility differences between other alternatives within her normalized utility function. This is the *normalize and average* procedure.

Some *normative* assumptions, which I will discuss in detail in section 5.4, may make it desirable that individuals with greater *overall* preference stability have more

[17]Remember that more stable stochastic preferences have P-values that are further away from ½ than less stable stochastic preferences.

φ	u_1	t_1	u_2	t_2	u_3	t_3
x	1.80	1.31	-4.83	-1.97	-0.08	-0.26
y	1.00	0.73	1.27	0.52	-0.39	-1.34
z	0.47	0.34	1.64	0.67	-0.23	-0.79
a	-0.19	-0.14	0.61	0.25	0.23	0.79
b	-1.23	-0.90	-0.15	-0.06	0.39	1.34
c	-1.84	-1.34	1.46	0.59	0.08	0.26

Figure 5.9: Individual utility

φ	avg(U^*)	avg(U_t)
x	-1.04	-0.31
y	0.62	-0.03
z	0.63	0.07
a	0.22	0.30
b	-0.33	0.13
c	-0.10	-0.16

Figure 5.10: Social utility

overall influence on the social choice function. To accommodate such a condition, the transformation function could be altered such that it only *aligns* individual utility functions to some common zero-point but does not normalize them. Specifically, set the zero-point of each individual utility function to the mean of that individual's set of utility assignments and translate all other utility assignments accordingly. Then the *translated* utility value for $u_1(x)$ is found through the following function:

$$u_1^*(x) = u_1(x) - \bar{F}_1 = 1.80 - .01 = 1.79$$

I do not show the complete translated values for all individuals and alternatives, because the translation is much simpler than the normalization function and so the reader may easily calculate them. By using such *translated* utility functions, preferences of individuals are weighed differently *on a whole* based upon their overall preference stabilities. This is the *translate and average* procedure. The social utility function using this sort of transformation is listed in figure 5.10, column *avg*(U^*).

These two transformation methods produce different strict social orderings of the alternatives in F, as shown below:

$avg(U^*)$: $z \succ y \succ a \succ c \succ b \succ x$ $avg(U_t)$: $a \succ b \succ z \succ y \succ c \succ x$

Translate and average ($avg(U^*)$) allows some small group of individuals — or in this case, a single individual — to dominate (in the non-technical sense) the social ordering. In the above example, the strict order of $avg(U^*)$ is largely determined by u_2. *Normalize and average* ($avg(U_t)$) is more *democratic* in this respect, but it still allows individuals with very stable preferences over some alternatives to have more influence on the social utility values of those alternatives. For instance, the very low utility value of x and high value of c assigned by I_2 significantly decrease the social utility assignments over those alternatives by $avg(U_t)$. I will discuss the implications of different *normative* assumptions on the types of allowable utility transformations next.

5.4 Normative assumptions

In the last section, I calculated cardinal social utilities using two different transformation methods and showed that the transformation one uses can significantly alter the resultant social utilities. Although the structural assumptions I provided in section 5.2 form the basis of certain interpretations of utility values, and subsequent comparisons between individuals, they do not themselves explain why certain differences in utility values might be *relevant* to social decisions. Such an explanation requires the acceptance of a set of *normative* assumptions about the significance of utility differences to the social decisions. This is necessary to legit-

imize the selection of a transformation method. The precise normative assumptions that should be made will be a matter of some controversy, similar to problems in political philosophy dealing with issues of fairness or the choice of a voting method in Arrovian social choice theory. In the case of social choice theory, different voting methods satisfy different sets of Arrow's conditions on democratic choice, though it is impossible that they all be satisfied. The choice of a voting method within any given voting context will typically depend upon which conditions are appropriate within *that specific context*; for instance, it may be appropriate to give each person veto power over some alternative pair in a committee meeting, but not in a national presidential election. These questions, unlike the aspects of my model that I presented in the previous two sections, are not technical in nature. And although formal proofs might show the logical consequences of adopting some set of conditions on a voting theory, they cannot answer the basic question of which conditions should be satisfied.

I have utilized a social choice method that *normalized* and then averaged individual utility scales. This method is responsive to the relative stability of preferences *within* each individual preference set, but is not responsive to overall differences in preference stability *between* individuals. I also introduced a second social choice method that merely *translated* individual utility scales, such that they were aligned, before finding average social utility values. This method is sensitive to both differences in stability within individual preference sets as well as *overall* differences between individuals. These two methods rely upon different, though related, normative assumptions about what constitutes a good social choice method. *translate and*

average relies upon *Expert Preference* whereas *normalize and average* relies upon *Special Interests* and *Overall Fairness*, which I introduced in section 5.2.

Any use of public opinion that instantiates the general will must satisfy something like the *Overall Fairness* condition. A condition of basic equality exists between citizens within the deliberative setting that helps constitute the *general will*: the arguments of all are given weight only according to the merits of their content and in eventual decisions all individuals are given equal vote. Any violation of *Overall Fairness* requires the representative to judge the merits of an individual's opinion, determining definitively that the opinions of one individual are more worthy of consideration than the opinions of others. For instance, a representative might give greater weight (in the fashion of *translate and average*) to those individuals who exhibit more stable opinions. But given that the measured patterns of individual opinions leave ambiguous the precise psychological process that brought about those opinions, any attempt to make the more detailed interpretation of opinion information would require that seemingly inadmissible information be brought into the analysis. Referencing the attitude model I introduced in figure 4.2 on page 116 will aid in seeing this. Systematic differences in the stabilities of different individuals' opinions arise in two primary ways. First, individuals may differ in how many considerations they generally entertain about the issues. Given similar probability distributions that determine which set of considerations are brought under consideration in any given instance of choice and similar proportions of considerations in favor of each alternative, a larger set of considerations will result in more stable responses. Some may be tempted to say that those who come to decisions after

entertaining more considerations should have more overall influence on social decisions. However, given the opinion data alone, it is impossible to determine whether two individuals differ in the number of considerations they entertain or rather the distribution over those considerations. It might be possible to independently determine the number of considerations an individual is utilizing in her decision but I think that even if this could be done, meaning that the precise structure of her attitude is *transparent*, it would not be consistent with the general will. This requires that the opinions of some individuals are favored over others, based upon judgments by representatives about the general merit of those *individuals* as sources of opinion.

It is instructive to see how *Special Interests* may be compatible with the general will in ways that *Expert Preference* is not. The core feature of *Special Interests* is its responsiveness to opinion strength relative to an individual's pattern of responses. By referencing the deliberative interpretation of the general will that I have considered here, it can be seen why the *general will* should be responsive in this manner. There are two primary ways in which greater opinion strength would seem to matter in a deliberative setting. First, it may signal to others a greater degree of expertise or consideration of an issue, especially if it appears that the strength of opinion has arisen from a resilience of the opinion to factors of the questioning not relevant to the issue. If those "expert" respondents direct their policy judgments at what is in the common good, as have said public opinion responses should, then their preferences will tend to better track what is in the common good than the preferences of those with less expertise. However, as I described above, the source of stability is typically ambiguous, therefore it must also be appropriate to use more stable opinions that

arise from stronger opinions. This is relevant to the general will in its effect on what the common good truly is, at which deliberation that determines the *general will* is directed. The satisfaction of a genuine concern about some issue seems to have some relevance to what is in the common good, because it is (all things considered) better for people be satisfied than for them not to be. In the deliberative setting, it would seem that the expression of great concern from some minority, even a small one, should have a significant effect on the resultant social decision. Therefore, it is reasonable that greater concern of some subgroups is reflected in the social decision. I have not intended this to be a complete defense of *normalize and average*, but only the basic outline of how such a defense might go. However, I do think that this outline provides reason to think that a reasonable normative argument could be given for the acceptance of this decision method, or something like it.

This finishes the main argument that I began in chapter 3: the model I have presented in this dissertation shows how political representatives can balance measures of public opinion with those of their constituents' interests. In the next chapter, I will show how the model of cardinal public opinion can be integrated into the decision model I introduced in figure 3.1, on pg. 68. I also address some implications of the model I have presented, particularly for public opinion survey methodology, deliberative democratic theory, and social choice theory.

Chapter 6: A Complete Model of Policy Decision

I have presented a model of policy choice by public officials that demonstrates how they can act according to their duties to respect public opinion. I argued that this duty has particular significance for political representatives, who are a special class of political officials tasked with integrating the will of the people into public decisions. In the last two sections, I demonstrated that a coherent concept of public opinion can be developed, and that it is also possible to measure it. But this still leaves plenty of questions about how exactly political representatives should act given that representative democracies require decisions to be made through deliberation. Governance is a complex process, and it is difficult to see immediately how the decision-theoretic framework I have developed can be integrated into a more realistic normative theory of governance. Such a normative theory should provide representatives with a better understanding of how they should act, if they are to act in accordance with their duties.

In this chapter, I will present a larger picture of how public opinion might be integrated into the decisions of political representatives, and how that integration fits into the systems of governance that are common in modern democracies. This picture will present a number of implications for modern democratic theory. This

is a formal model of an informal process that representatives use to decide what policy alternatives to pursue. The actual *process* they use to decide how to act will be far less precise and is unlikely to resemble the decision model itself. Rather, the model is useful in determining whether the decision process used by a representative diverges significantly from what is required for a representative to respect public opinion. First, I will present the "balancing procedure" in a slightly more complex form, showing how different criteria of policy choice might be integrated into an all-things-considered choice. I think that although the details of how different criteria are integrated may be difficult to determine, the basic structure of such an integration seems rather unproblematic. However, a far more puzzling problem emerges when these *aggregative* criteria are integrated into a *deliberative* decision process. Modern representative democracies are typically characterized by being deliberative, which is often thought to be vital for the rationality of public decision making.[1] This produces a conflict between two types of decision processes: those that are aggregative and those that are deliberative. I take the distinction between these decision methods to be one of the central rifts in democratic theory; in chapter 2, I argued that both of these types of decision procedures are needed in an instantiation of the general will, but it is not obvious how they might be meaningfully integrated into a single decision procedure. Here I begin an answer to this question, for which I nonetheless do not think that I have an entirely adequate response. I argue that this is a problem for democratic theory in general and not just for the view I have presented here, and that the precision with which I have presented the

[1]See Pettit (1997) and Cohen (1989).

aggregative decision process through the balancing procedure gives a clearer picture of the basic conflict between aggregation and deliberation. This will become clearer after the consideration of an example.

6.1 The weighing procedure

Consider the policy choice example I began in section 4.4. Some political representative has a constituency containing three individuals or groups $\{I_1, I_2, I_3\}$, individuated by their opinions toward a set of public policies under consideration F_i:

x Nationalized health care system

y Free health clinics for the poor

z Subsidies for full private health insurance for the poor

a Subsidies for catastrophic health insurance for the poor

b More tax deductions for health care costs

c Status quo

The aggregate opinion of this constituency can be represented as the cardinal utility function obtained using *normalize and average* from figure 5.10, pg. 167. I reproduce this utility function below:

φ	x	y	z	a	b	c
avg(U_t)	-0.31	-0.03	0.07	0.30	0.13	-0.16

One of the purposes for finding such a function is to permit the balancing between a quantification of opinion and a measure of welfare (or the common good). But it is not entirely straightfoward how such a comparison might be made.

'Welfare' itself is likely composed of different components, and therefore formulating an appropriate welfare measure will be no simple matter. And "welfare" is just something I am using to stand in for the concept of the "common good." It is reasonable to think that representatives with different political ideologies would conceive of welfare in different ways, and so I will first consider how two representatives with different conceptions of welfare and/or different estimations of the effects on welfare of various policies might use public opinion of identical constituencies in their policy decision making. Consider two representatives R_C and R_L, who have ideologies that might best be described as "American Conservative" and "American Liberal" respectively. Assume that each individual assigns estimations of "welfare" that will be produced by each policy, shown in figure 6.1; it will be initially unclear what these values might stand for, but I will explain this further in a subsequent example.

The values of the numerical representations in R_C and R_L are symetrical with one another, and it may be possible to infer significance of the 0-point, such that negative values represent expected "losses" and positive values represent expected

φ	R_C	R_L	R_C^*	R_L^*	avg$(U_t)^*$
x	-1	1	-0.55	0.55	-0.56
y	-0.8	0.8	-0.35	0.35	-0.05
z	-0.9	0.9	-0.45	0.45	0.13
a	-0.2	0.2	0.25	-0.25	0.54
b	0.1	-0.1	0.55	-0.55	0.23
c	0	0	0.45	-0.45	-0.29

Figure 6.1: Welfare estimations by R_C and R_L

"gains." However, I have not presented any underlying theory for how measurements of welfare might be given this significance, so it is unlikely at this point that such an interpretation could be justified. The model does not require that welfare estimates be meaningfully compared between *representatives*, but it does require that there is a meaningful way to make comparisons between each individual's estimates of welfare and the measurement of public opinion; without such comparisons, the balancing procedure cannot make use of cardinal information. This would make the results of the balancing procedure arbitrary. One simple way to make such a comparision is to match the *ranges* of the scales that are to be compared. Without any theoretical rationale for interpreting criteria in a particular way, there is no basis for any transformations beyond simple features of the data itself. For instance, in figure 6.1 avg$(U_t)^*$ is a transformation of the results of the *normalize and average* procedure (from figure 5.10 on pg. 167) that matches the ranges of R_L and R_C. R_C^* and R_L^* are then translations of those measures of welfare such that their means match the mean public opinion value found in avg$(U_t)^*$. This renders R_C^* and R_L^* comparable to avg$(U_t)^*$ using only basic statistical information of the scales, and

not any substantive assumptions about the sources of the scales.[2]

With simplifying assumptions in place, one can see what a weighing procedure might look like, and how the estimates of welfare might change the policy a representative should pursue. For instance, if equal weights are applied to opinion and welfare, then the utility function that results from an averaging balancing procedure can be seen in columns R_C^* and R_L^*, subcolumn 1 of figure 6.2.

	R_C^*			R_L^*		
Weight on welfare $\rightarrow$	2	1	1/4	2	1	1/4
x	-0.83	-0.55	-0.35	0.27	0.00	-0.21
y	-0.38	-0.20	-0.07	0.32	0.15	0.02
z	-0.39	-0.16	0.01	0.51	0.29	0.12
a	0.52	0.40	0.30	0.02	0.15	0.24
b	0.66	0.39	0.19	-0.43	-0.16	0.05
c	0.30	0.08	-0.09	-0.59	-0.40	-0.20

Figure 6.2: All things considered utilities for R_C^* and R_L^*

A weighted average can represent a balancing procedure that weighs one criterion greater than another. There are many reasons why a representative might apply specific weights to public opinion and her estimates of welfare; some of these reasons may be stable while others may vary depending upon the issue at hand. For instance, representatives may give more weight to public opinion for policy issues that are moral at their core than those that seem to require a deep understanding of economics. This can obviously have an effect on the way in which representatives ultimately rank the alternatives. For instance, the following different rankings result from the different weights on estimates of welfare:

[2]I do not think that that I have given anywhere near an adequate justification for the use of this procedure. However, I will argue below that this problem is a liability for *any* (cardinal) model of multi-criteria choice.

R^*_C

$$2 \;\; b \succ a \succ c \succ y \succ z \succ x$$

$$1 \;\; a \succ b \succ c \succ z \sim y \succ x$$

$$1/4 \;\; a \succ b \succ z \succ y \succ c \succ x$$

R^*_L

$$2 \;\; z \succ y \succ x \succ a \succ b \succ c$$

$$1 \;\; z \succ y \succ a \succ x \sim b \succ c$$

$$1/4 \;\; a \succ z \succ b \succ y \succ c \succ x$$

It is not entirely clear what should be made of such formal representations. Every element seems rather arbitrary, from the utility values that represent welfare to the weights of welfare against public opinion. The assessment by representatives of the welfare *ordering* of alternatives is likely not arbitrary, but it is unclear how the cardinal values themselves are produced. If I am to provide even a rough outline of a decision model by representatives then I need to say something about how cardinal utility values for welfare might be found. I have already presented welfare as the element of a complete choice model in figure 3.1 on pg. 68. Figure 6.3 (pg. 180) shows an example of the expected utility involving the set of alternatives F_i.

The numerical values of outcomes (o_n) could stand for many different quantities relevant to welfare; for instance, thousands of lives or dollars saved over the status quo. Or these outcomes could be composite values derived from a set of

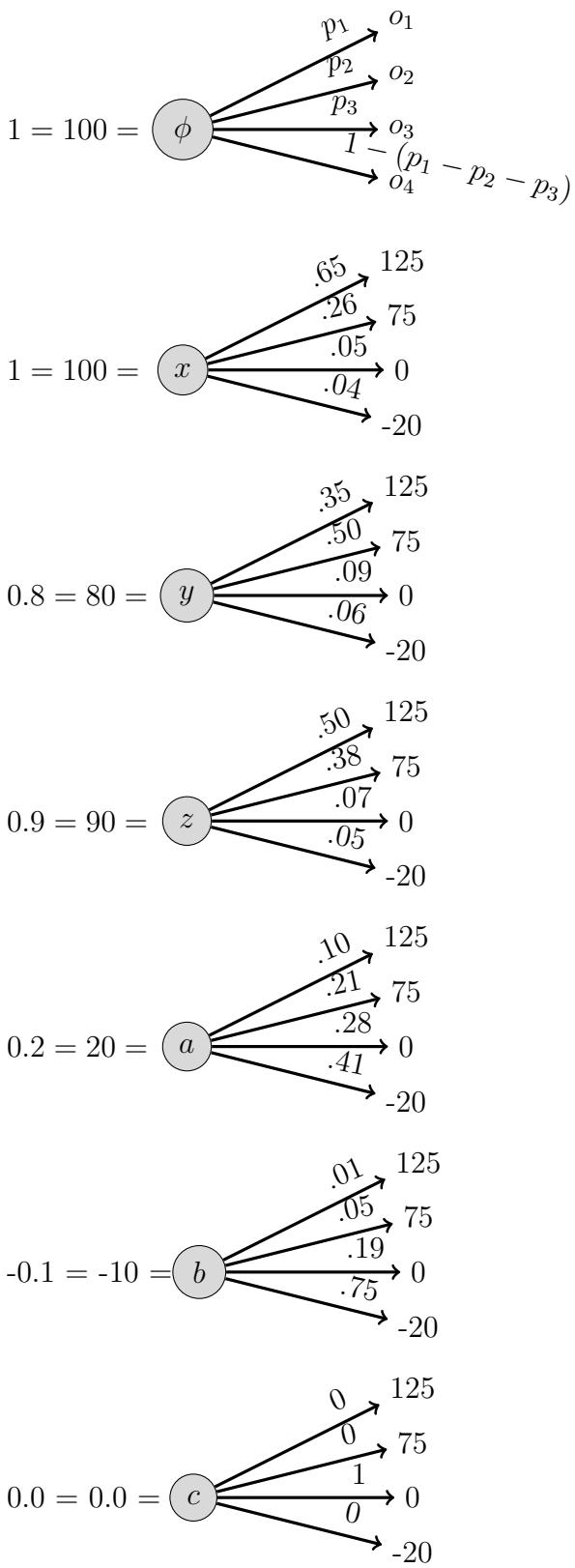

Figure 6.3: Policy choice model

welfare criteria. However the outcomes are determined, they are then multiplied by their probabilities to find the expected utility of each alternative policy. What I give here is a simplistic example, but in the full model the utilities for outcomes are multiplied by the probability distribution over those outcomes to find the expected utility.[3] I will presume that representatives should use some standard measure of expected welfare in their weighing of their estimates of welfare against public opinion, though this may not be the case. There is some reason why one might, for instance, prefer a maximin rule or one that weighs negative outcomes greater than positive ones.

In the example I give in figure 6.3, a scale that seems naturally comparable to the public opinion scale can be easily derived from the values; dividing the expected values by 100 produces a scale in the -0.1:1 interval, which is similar to the interval containing the public opinion scale. Because a weight is simply the multiplication of some scale value, if a representative is to respect public opinion by weighing it roughly equal to her estimates of welfare then she must also transform any measurement of welfare such that it is on a similar scale as public opinion. Again, this process seems quite arbitrary and perhaps poorly describes how representatives likely make decisions. Certainly a complete model of the balancing procedure would require a better picture of how this can be done non-arbitrarily. It is not within the scope of this dissertation to provide such a model. However, I there is good reason to think that something like this model can explain disagreement on normative

[3]It should be noted that finding the expected utility of some choice is far from trivial, and how precisely this should be done is likely to be controversial. For instance, one could have different views about how future utility should be discounted, or whether it should be discounted at all.

issues in many different types of multidimensional choice. For instance, if officials are debating about what sort of bridge should be built across a river, there may be several dimensions relevant to the choice. The first natural step in such a decision will be to reduce as many factors to purely monetary values as possible. This might work for the effects of a bridge on business as well as its initial and recurring costs. However, aesthetic and environmental impacts of the bridge, along with effects like commute times might only be completely reduced to monetary values with great difficulty and at the risk of losing contact with what is being measured. Therefore, such non-monetary criteria must be part of the decision, though likely not part of any formal cost-benefit analysis.[4] It makes sense for someone to say that, for instance, the impact of a particular bridge on the aesthetic appeal of a city is not being given enough weight in a public decision process. The fact that a bridge is ugly has weight in a decision of whether to build it beyond the monetary impacts of that ugliness, but such a fact can be outweighed by the fact that the bridge will be exceptionally beneficial to the local economy. And it would be unreasonable to think that such a weighing procedure should be impacted by the economic *unit*, whether they be pesos or dollars. It is similarly clear that decisions about such weights have real impact on the overall decision. The weights given to aesthetics by officials in Paris will likely be greater than those by officials in Houston, and this has a significant impact on the way decisions are made. The decision of what weights should be used when determining the overall impact on welfare of some

[4]In fact, such criteria are often included in cost-benefit analyses through monetary proxies, such as impacts on tourism and destruction of monetarily quantified environmental resources, but these proxies do not *exhaust* the inclusion of these criteria in the decision.

policy is primarily normative, but one that certainly plays a role in public decisions that involve composite measures of public welfare.

Returning to the example, it can be seen how a decision process modeled after the balancing procedure might function using a cardinal measure of public opinion, cardinal values of outcomes, and probability distributions over all of the outcomes for each policy alternative. The result of balancing can be best seen when there is some conflict between the representative and public opinion. Consider R_L, and its translated form R_L^* from 6.2. When this is balanced against the translated measure of public opinion ($\text{avg}(U_t)^*$), the resulting ordering is shifted from the representative's estimates of welfare, toward the measurement of public opinion. This is especially obvious in the case of the representative's ranking of z. In both of the representative's welfare estimates and the measure of public opinion, z is ranked near the middle, but it is elevated to the highest ranked alternative when those are balanced against one another. A closer examination of the cardinal values in these rankings reveals that R_L assigns a relatively high value to z, and the public opinion assigns it a value just above that scale's mean value. However, there is a quite a bit of conflict about the values of y and a, and this results in these alternatives being ranked below z. This simple averaging procedure is just one way of balancing public opinion with representative estimates of welfare, and there may be advantages to other types of procedures.

Of course, if representatives are free to choose any weights, then it is possible for the representative's own welfare estimates to dictate what policies she pursues; clearly something must be said about what sorts of *weight* representatives must give

public opinion. I think a presumption of rough equality of weights is a natural starting point. It seems that this pays respect to the judgments of the public while at the same time recognizing that the representative is in an elevated position of judgment in relation to the public. Though this initial weighing is simply a presumption, and can be adjusted depending upon the representative's assessment of the competence and reasonableness with which members of the public made their judgments, any adjustments to the weights must be reasonably justifiable to the public. This, I think, arises quite naturally from the respect that must be given to public opinion.

6.2 Implications

With the policy decision model completed, it is easier to understand some implications of the model. Consideration of these implications will also reveal the complications for any public decision procedure of this sort that I have presented. These implications and complications fall into three primary categories: the measurement of public opinion, the integration of aggregative and deliberative decision methods, and social choice theoretic models of phenomena other than public opinion. In the rest of this chapter I will consider each of these in turn.

6.2.1 The realistic measurements of public opinion

In this dissertation I have provided a model of public opinion. I have described this model in measurement theoretic terms, though I must now reiterate that the

subject matter of this dissertation has *not* been a method of measuring public opinion; this is not a work of survey methodology. Rather, what I have presented is model of what the phenomenon of public opinion is, and especially its characteristics that might be relevant to its measurement and use by political representatives. This is meant as a response to the view that no public opinion relevant to policy choice exists. The provision of such a model of public opinion is important because it shows that public opinion could possibly have a significant role in the decisions by political representatives in democracies. Although the fact that public opinion *could* be measured is important to its structure, I do not intend the model itself to be taken as a description of a measurement method.

According to the model, a direct measurement of public opinion would require that every individual in the population be asked to compare every alternative over some large number of trials. For large populations this requirement is almost certainly impossible to satisfy, at least if it is assumed that most human activities are not devoted to measuring public opinion. The requirement that attitudes be formed by repeated pair-wise comparisons of all alternatives is also likely impractical in many cases. Although similar measurement methods with large numbers of trials are common in the pyschophysical research, especially in the study of the human perceptual system, those studies typically require very few participants (often fewer than 20) to adequately study their phenomena of interest.[5] It may, however, be possible to approximate this direct measure of public opinion if other simplifying

[5]For an idea of typical sample sizes in such psychophysical research, see some recent articles in *The Journal of Vision*, which almost exclusively publishes research that use these standard psychophysical methods.

assumptions are made. One set of assumptions is necessary to specify the relationship between a population and a *sample* of that population. Such a *sampling* model is necessary for pretty much all measurement in science, though it requires certain assumptions about the basic structure of opinions in the population, which are not required in standard social choice models. I will argue that such a model can probably be applied to public opinion measurements. An additional strategy for making public opinion measurements is to *estimate* a measurement of cardinal public opinion without employing a cardinal scaling technique on each individual attitude set. One set of methods, used by Thurstone in his early studies of public opinion, collect cardinal public opinion information directly without first collecting *individual* cardinal attitude information. I think that these methods have promise for providing a cardinal measure of public opinion, though they are unable to provide information about *individual differences* in opinion strength. A second type of method uses purely ordinal measures of public opinion, such as the Likert scale, plus assumptions about the distribution of opinions in the public to infer cardinal public opinion information. I think that these additional assumptions are typically implausible, and so I do not think that such Likert-type measurement methods show promise in providing cardinal public opinion information. As I will describe, I am unsure about the plausibility of any of the assumptions behind these simplification methods. However, I think that this is one important question in need of empirical investigation. If my arguments in this dissertation have been successful, then the ability of standard public opinion measures to inform political representatives of public opinion hinges on the ability of those measures to approximate cardinal

public opinion.

I will consider several simplification strategies for measuring public opinion, but first I must specify the sort of public opinion research that I will be concerned with. There are several distinct goals that public opinion polling can have. I will focus on three major goals: the prediction of election results, the discovery of individual characteristics that co-vary with opinion, and the content of public opinion itself. For the first two types of goals, the content of public opinion is merely instrumental to the true end of the research. A large proportion of polls are of these first two types. For instance, much of the opinion "tracking" polling done during election campaigns is attempting to predict the eventual election outcome.[6] One hallmark of such polls is the concern with the composition of the sample; because some individuals in the electorate are more likely to vote than others, a good prediction of the election outcomes relies upon an accurate prediction of who will vote. A poll that accurately captures that opinion of the entire electorate toward the candidates might nonetheless fail to predict the winner if those who vote are significantly different in their attitudes than those who do not vote. Although the most direct questions for the purposes of prediction ask something like "If the election were held today, would you vote for Barack Obama or Mitt Romney?", other questions might also be useful in prediction. For instance, if a poll shows that people approve of items in one candidate's platform more than another's platform, this may bode well for the first candidate's prospects. Such polls can also be used by candidates to decide what parts of their platforms to emphasize in a campaign or the content of the

[6]One prominent example is the Gallup tracking poll.

platform itself. Nonetheless, this is all done to maximize the candidate's chances of winning the election. If it is discovered by a candidate that those who are unlikely to vote prefer some policy alternative x whereas those who *are* likely to vote prefer policy y, then it would be prudent for the candidate to publicly advocate policy y. Those interested in public opinion for this purpose need not be concerned with representing public opinion in its full richness (i.e. cardinally) unless it is likely to aid in the prediction of election outcomes. And because I have not presented a theory to link public opinion and election behaviors, I have not provided those interested in predicting elections reason to prefer richer representations of public opinion.

Polling by academic researchers is aimed at another goal besides the prediction of elections. Rather, they are concerned with the determinants (or correlates) of certain types of attitudes. Public opinions are analyzed in the hope of understanding how and why people (and groups) come to have the attitudes they do. A great deal of important work in political science is of this sort. For instance, Converse's use of public opinion in "The nature of belief systems in mass publics" was limited to general facts about public opinion, and specifically its instability for most types of policy questions. A more recent example is Andrew Berinsky's *In Time of War*, which investigates the common patterns of public opinion formation during several American wars. Berinsky shows that the party identification and political knowledge of individuals are highly correlated with public opinion. Others have found that individual policy opinions are greatly affected by individual demographics, such as race, sex, class, and income.[7] This work is not directed at finding the content of

[7]For useful summaries of some of this literature see *The Oxford Handbook of Political Behavior*

public opinion itself at any given moment in time or for any specific constituency, but rather only general patterns of public opinion. Similarly to the predictive use of public opinion, I have not provided good reason for such researchers to prefer cardinal representations of public opinion.

I am concerned with a third use of public opinion polling that attempts to measure the content of public opinion at particular times and in particular constituencies. More specifically, I am interested in its use in decisions by political representatives. As I have argued in this dissertation, public opinion measurements with this as an aim must take into account the cardinal structure of public opinion in order to be adequate. The next question is whether it might be possible to obtain cardinal information about public opinion using methods more easily implemented than the idealized Thurstonian method I have presented. I will first consider assumptions that would permit the use of sampling methods to measure the public opinion of some population. I think these assumptions are plausible, and are required by any implementable method of public opinion measurement. I will then consider a number of methods that further simplify measurement by relying upon assumptions about the structure of public opinion (collective attitudes) in societies, rather than the assumptions about *individual* attitudes that I have made in my own model. As I have already said, I am less sure about the veracity of these methods.

I will first consider whether the opinion from a random sample of a constituency can be used to measure the public opinion of the entire constituency. This is crucial for any practical measurement method of public opinion, because

(2007), chapters 24-27.

it is almost never possible to measure the attitudes of all members of the public. However, if one measures a large enough *random sample* of the public, then the measured opinion of the sample should approximate that of the public. This is based upon very general statistical principles. If a random variable is distributed normally throughout a population, then the mean of a random sample approximates the mean of the population, and converges to that mean as the size of the sample increases. However, public opinion measurements may not satisfy the necessary statistical assumptions. Without a theory showing why *actual* opinions within the public will tend to be normally distributed there is little reason to think that the measured opinion of a sample will adequately approximate the entire public opinion. However, this statistical justification for sampling is often subordinated to a *normative* justification. Random sampling in survey methodology has often been said to give all individuals an equal opportunity of being included within the sample, though the measurement of the opinion of any given sample is not guaranteed to approximate the true public opinion (that of the entire population).[8] Such a view contends that a sampling procedure is *legitimated* by being part of a *democratic* procedure rather than because of its ability to accurately *measure* public opinion

The "democratic" justification for sampling in public opinion surveys has been prevalent since their earliest uses, but it is a far less natural justification for sampling than one based upon accurate measurement.[9] Sampling for the *measurement* of public opinion must have the goal of accurate measurement rather than being

[8]See Althaus (2003), chapter 7 for a discussion of sampling in public opinion survey methodology.

[9]Perhaps one of the earliest examples of the "democratic" justification for sampling is given by George Gallup (1940).

an appropriately democratic procedure, because otherwise the *content* of the *entire* public's policy attitudes may not be reflected in the representative's decisions. Only the members of the sample would be properly represented. Therefore, I think that only an *accurate* measure of public opinion is appropriate to serve the role in public decisions that I have described in this dissertation. The problem, given the likelihood that individual attitudes are not normally distributed in populations, is to determine what makes some sample of individuals a "good" sample with respect to public opinion. One common method is to construct a "representative" sample of the public by matching certain *demographic* (socio-economic status, race, sex, etc.) features of the sample to those of the public, and in so doing *descriptively represent* the population within the sample. Descriptive representation, unlike random sampling, specifies the features that individuals within a sample should possess. Some sample is *descriptively representative* if it matches some general description of the population.[10] Although in a trivial sense this description could just be the aggregate opinion of the population, that is an unhelpful description when the public opinion itself is unknown. Rather, the relevant features within the population must already be known and tend to co-vary with public opinion in order for them to play any role in the evaluation of the adequacy of the sample. A match between the sample and the population with respect to these descriptive characteristics may then bode well for a match with respect to public opinion.

It is well known that certain demographic characteristics are correlated with

[10]See Pitkin (1967), chapter 4 for an influential discussion of descriptive representation.

individual attitudes about many policy issues.[11] Such demographic characteristics of a sample are often used to make adjustments to results of public opinion surveys in order to estimate the results that would be obtained from a sample well matched to a population's demographic characteristics. These characteristics provide one way of defining how a sample should descriptively represent the population. However, one might worry that given the modest character of these correlations, it is not obvious that demographic representativeness can *guarantee* that a sample's opinion will approximate the opinion of the entire population. I think this worry sets the standard too high. Public opinion surveys need only approximate the public opinion substantially better than the other available methods. And although methods such as personal communication with constituents and town hall meetings have long been used by public officials to understand public opinion, samples formed using these methods are well-known to substantially differ in both their demographic characteristics and policy opinions from those of the populations they supposedly represent (Verba, 1995). Such bias is even more apparent in the sample of the electorate that chooses to vote in any given election; in fact, the nature of this bias is sometimes crucial in determining many election outcomes.[12] At issue is whether public opinion surveys using samples can make important contributions to public decision making, and I think that they clearly can. Such samples descriptively represent the population better than other methods, and so it would seem that public opinion surveys using samples can improve representatives' understanding of

[11] See Page and Shapiro (1992), chapter 7 for a summary of some of this research.

[12] The bias introduced by voter turnout is an important and constantly changing factor in the determination of election outcomes. For example, voter turnout in the 2008 and 2012 presidential elections likely had a major impact on the results (Frey, 2013).

the content of public opinion.

My argument then can be stated as follows: a sample's aggregate opinion is capable of approximating the public opinion insofar as the sample descriptively represents the population in the ways that are relevant to attitude formation. Additionally, if such a sample is able to approximate public opinion better than other standard methods, then sampling methods are useful tools for the measurement of public opinion. I do not think that I have adequately defended this argument here; it requires that several empirical assumptions be true, and I have not provided enough empirical evidence to know whether this is the case. However, I have suggested how the aggregate opinion of a sample might adequately measure the public opinion of an entire population.

The possibility that sampling methods might approximate a measurement of public opinion does not eliminate the demandingness of the measurement task I have presented here. Each individual in the sample must provide pair-wise choices for each alternative pair numerous times. In order to make it impossible for individuals to remember and repeat past responses, this method will likely require that the individuals are asked about each pair using many different questions or during several sessions. Although possible to implement, such a method would be costly. Also, it may be unnecessary; in his studies of attitudes, Thurstone himself typically used a simplified method that only requires each individual to make a single comparison for each alternative pair.[13] Using this method the aggregate attitude measure is

[13]See Thurstone (1959), chapter 7 for an example of finding public attitude scales for the perceived seriousness of crimes and a related scaling method in chapter 11 used to find public attitudes toward various ethnic groups.

formed by finding the percentages of choices (P-values) for each alternative in each pair *across* individuals and then producing a scale from those percentages in much the same way as I did in section 4.3 with *individual* responses. Each individual is only asked to compare each alternative once, and so the demands on individuals are greatly reduced. It also eliminates the need to formulate many different questions that compare the same two alternatives, because there is no risk of a single individual remembering and repeating her responses to identical comparisons. However, this changes the subject of the measurement. In my model, individual attitudes are measured and then aggregated to form a public opinion. When using the simplified method, Thurstone measured public opinion directly by obtaining pair-wise comparisons from a sample of the population. If it is assumed that each individual evaluation of an alternative (in terms of the utility value that determines their pair-wise choices) is a sample from the population of evaluations distributed about a mean, then the mean evaluation of that sample approximates the evaluation by the population — this is the public opinion. In statistics, such a measurement method is said to use a *between*-subjects design, whereas the model I have presented uses a *within*-subjects design.[14] The advantage of a *within*-subjects design is its ability to utilize information about each individual's response (or error) tendencies in the analysis of responses. For a between-subjects design, individual information cannot be used in this manner.

For the between-subjects variant of Thurstone's method, there is no means

[14]I provide a short discussion of the differences between between-subjects and within-subjects designs in section 5.1, starting in the paragraph before figure 5.1, pg. 145.

of representing individual or group differences of attitude strength. In fact, if such differences are pervasive in the population, it may undermine the analysis entirely. However, one apparent benefit of this method is the way it avoids explicit interpersonal comparisons of attitude strengths altogether. The only measurements that are compared across individuals are their responses to pair-wise choices between alternatives. The proportion of choices are then calculated for the entire population and scale values are calculated from them. This ignores all information about differences in individual attitude strength, but it may provide an important *estimate* of cardinal *public* opinion, especially if individual attitudes are well structured in the population. Although such an assumption is required, it does offer substantial benefit over the purely ordinal methods of measuring public opinion.

The between-subjects Thurstonian method is still more involved than what most attitude researchers would find practical. But there are a number of other psychophysical techniques that may be used to make estimations of Thurstonian scale values.[15] And it should be noted that Thurstone most often made use of such estimates for his studies of attitudes; although he considered the pair-wise method to be the gold-standard that he was approximating, the practical requirements of measurement often called for only estimates of scale values. Probably the most important of these methods is the *Method of Equal-Appearing Intervals*. This requires the construction of a cardinal measurement scale by a set of "judges" who evaluate a set of evaluative terms according to their favorability, with each attempting to

[15]For an introduction to some of these techniques see Eagly and Chaiken (1993), chapter 2, especially pp. 30-44.

create a scale with equal scale intervals between all terms. Scale values for all these terms are then calculated from these estimates, and a measurement scale is formed using only terms that were judged as being separated by equal intervals.[16] This scale is then used by subjects (who are distinct from the judges) in the sample to rate the items of interest. Judges are responsible for providing cardinal scale values for the terms used by the sample to evaluate the alternatives, and the subjects evaluate the alternatives using those terms. For instance, if four terms ('unhappy', 'indifferent','support', and 'overjoyed') are used to evaluate some set of policy alternatives x, y, z, then judges first assign integer values to those terms from some predetermined range and the subjects assign one of the terms to each policy alternative (perhaps 'unhappy' toward x, 'indifferent' toward y and z). When a larger number of terms are rated in this way, it permits a range of cardinal values to be represented by qualitative terms in opinion surveys. If it is assumed that the value assignments of the judges have cardinal significance, and the subjects use those terms in a similar fashion as the judges, then this should provide cardinal assignments of value to the policies.

It is unclear what to think about the *Method of Equal-Appearing Intervals*. In this method, judges attempt to make equal intervals between the terms, and this is supposed to produce a scale of value for the terms that *does* have equal intervals between the terms. This requires that the judges themselves be capable of estimating the equal intervals, and for scales to be comparable across judges;

[16]Because scale values assigned by the judges will differ, the median scale value from the judges' value assignments is used for each term in the analysis of opinion.

or at least it requires that differences in the ways in which judges make estimates of these intervals be normally distributed. It also requires that the intervals of value assigned to the terms by the judges match those that would be assigned by the subjects. Only then can the process of judgment be capable of producing scale values for the terms as they are used. None of these assumptions are required by the method of direct pair-wise comparisons of the *alternatives*. When a cardinal scale is derived from a set of pair-wise choices, there is no need for any individual to access her own estimates of scale intervals; those intervals are derived from behaviors alone plus a stochastic model of how those behaviors are produced from a set of attitudes toward the alternatives. Some concern of subjectivity can be eliminated by use of the *Method of Successive Intervals* (rather than *Equal-Appearing Intervals*), which finds scale values of the terms from the *proportions* of judges rating the various alternatives either better than or worse than the other terms. Therefore, judges need not make numerical assignments of value to the terms. This only requires that judges form *orders* of the terms according to their approval of them and does not assume that judges are able to place terms along a scale of equal intervals. Rather it finds the intervals from the attempts of judges to place terms along such a scale. This then eliminates the requirement that judges are able to correctly produce scales of equal intervals, but does not eliminate the need for the judges' estimates of values to match those that would be obtained from the population. If there is a great deal of agreement in society about the value associated with different evaluative terms then this method is promising, but it is unclear whether that is generally the case.

I have said nothing in this section so far about the *Likert Scale*, which is to-

day by far the most common method of scaling survey items. Its dominance spans multiple disciplines, from personality tests to public opinion surveys. And yet in section 3.3.3 I argued that Likert scales offer only ordinal information about preferences and therefore are of little use to political representatives. This is actually a matter of great controversy; Renis Likert himself thought that his scale could replace Thurstone's techniques and produce an estimate of a cardinal attitude scale.[17] The Likert scale avoids the need for judges to form an interval scale from the set of evaluative terms, because subjects rate each alternative directly from the scale the judges use in the Thurstonian method to rate evaluative terms. Likert argued that his scaling method matches the reliability and validity of the Thurstonian methods, while being much simpler to implement. Although Likert's method seems to have won the day, there is evidence that it is not as suitable to replace the Thurstonian scaling method as Likert thought.[18] However, it is possible that Likert scales may be adequate regardless of whether they are able to *measure* cardinal public opinion; cardinal information may instead be *inferred* from ordinal opinion information. If the right assumptions are made about the distribution of attitudes in the public, then it may be possible to infer this cardinal information without using a method that measures the information directly.

Consider a set of policy alternatives x, y, z, each of which is assigned an integer scale value by each individual, with high numbers representing greater approval. So if x, y, z are assigned the values 5, 4, 2 respectively by some individual, then that

[17] This is the subject of Likert's 1932 PhD thesis, *A Technique for the Measurement of Attitudes*.
[18] For instance, see Roberts et al. (1999) and Drasgow et al. (2010).

individual ranks the alternatives $x \succ y \succ z$. This makes no assumption that the intervals between the values are significant. Next assume that the percentages of individuals making each numerical assignment are (some variant of) normally distributed about some mode value for each alternative. For example, the percentages in figure 6.4 may come from the aggregate of individual responses. In this example, the distributions of assignments of values are skewed-normal. Given that this is only ordinal information, it is not possible to find a *mean* value for each alternative, but an analysis of the distributions does provide some information that is relevant to the public rankings of the alternatives; the form that this analysis must take given the lack of cardinal information is instructive. The social ordering $x \succ y \succ z$ seems to be a reasonable inference from the data. This is supported by the observation that a majority of respondents rate x as better than 3 (80%), y better than 2 (70%), and z *worse* or equal to 2 (75%). This interpretation would be even better supported by unambiguous labels of the scale values (such as strongly approve, approve, indifferent, disapprove, strongly disapprove); such labels would provide a further basis for interpreting the results in support the ordering $x \succ y \succ z$. With such labels it seems reasonable to say that a majority approve (or strongly approve) of x, a majority is either indifferent to or disapproves of y, and a majority disapproves of z.

This reasoning runs into trouble if public opinion is taken to be an aggregate of *cardinal individual* attitudes, which is the view that I have advanced in this dissertation. The sort of response information found in figure 6.4 could originate from

$f(\varphi)$ \ φ	x	y	z
5	**30%**	15%	1%
4	25%	20%	4%
3	20%	**40%**	20%
2	15%	20%	**50%**
1	10%	5%	25%

Figure 6.4: Percentages of Likert scale value assignments

a number of different sets of individual cardinal attitudes. Consequently, we cannot definitively infer the structure of underlying cardinal attitudes. First consider attitude scale *differences* between the alternatives. From the distribution of percentages, it appears that $u(x) - u(y) > u(y) - u(z)$. A similar style of interpretation — and right now I want to emphasize the informal nature of these intuitive assessments — might yield an ordering on the utility *values* as well: $u(x) > u(y) > u(z)$. Both of these statements seem to intuitively follow from the information in figure 6.4, but they require substantive assumptions about how the aggregate opinion strength is connected to the percentage of individuals responding with a given value. More specifically, and described statistically, they assume that there is a clear relationship between the modes of the percentage information and the mean utility value for each alternative. There are plausible counterexamples to this assumption. Consider some minority group, which composes 15% of the total sample, with extremely strong positive attitudes toward y, just as strong negative attitudes towards z, and nearly as strong attitudes against x. That group's percentages are shown in figure 6.5. The percentages of the rest of the sample (85% of the total sample) are shown in figure 6.6.

$f(\varphi)$ \ φ	x	y	z
5	0%	**15%**	0%
4	0%	0%	0%
3	0%	0%	0%
2	5%	0%	0%
1	**10%**	0%	**15%**

Figure 6.5: Percentages of Likert scale value assignments by the Minority

$f(\varphi)$ \ φ	x	y	z
5	**30%**	0%	1%
4	25%	20%	4%
3	20%	**40%**	20%
2	10%	20%	**50%**
1	0%	5%	10%

Figure 6.6: Percentages of Likert scale value assignments by the Majority

Next assume that the minority's positive attitude toward y is far stronger than anyone's positive attitude (in the majority) toward x. Given that a measure of public opinion should take into account such differences in attitude strength, the real possibility of samples with complex opinion structures should call into question the statements about utility differences and values that I give above. The lack of information about differences in attitude strength should also further support skepticism about whether *any* information about attitude strength gleaned from Likert scales is significant. Different subgroups in the sample may use the Likert scale differently from other parts of the sample, and if those subgroups have different distributions of responses from the entire sample (as the minority group in figure 6.6), then a measurement of public opinion using only Likert scales would not represent the underlying cardinal attitude scale.

Lastly, I will consider a public opinion survey procedure that captures cardinal public opinion information by making use of a normalization procedure similar to what I used in section 5.2 to make individual cardinal attitude scales comparable. For instance, unbounded numerical assignments to the set of alternatives could be collected from each individual, which could be immediately taken as individual

cardinal attitude scales. Those scales could then be normalized and aggregated with one of the methods I introduced in this dissertation. Such a method would not require the pair-wise Thurstonian procedure, and as a result would be much simpler to collect. This is another way of asking about the precise value of the pair-wise method. As I have already argued in section 4.2, I think the pair-wise method is crucial for the proper measurement of cardinal individual attitudes. The pair-wise method only requires that alternative choices be based in some process that has a relatively simple form, similar to the attitude model shown in figure 4.4 (pg. 123). That model shows how an underlying cardinal attitude structure can produce a pattern of pair-wise choices. A *direct* measurement of the attitude scale would require that an individual is able to *consciously* access and reproduce the full cardinal structure of her attitude set. Though an individual may have conscious access to the full *ordinal* structure of her attitude scale, as well as the choice she would make between two alternatives, I find it highly unlikely that an individual would be able to consciously access and reproduce the full cardinal structure.[19]. It is not enough then to employ some normalization procedure to cardinally measure public opinion; the pair-wise scaling method for individual attitudes is also crucial.

I have intended this to be a short investigation of how a measure of public opinion might be estimated, even though a complete measurement of cardinal public opinion may never be possible for large populations. Given the demandingness of the scaling method I have presented in this dissertation, it is not suitable for the actual measurement of public opinion. Utilizing sampling procedures to estimate

[19]I provide a more detailed argument of this form in section 4.2.

the public opinion of a large population is likely unproblematic. However, between-subjects Thurstonian methods may only provide adequate estimations of cardinal public opinion in populations with relatively normally distributed opinion profiles. Standard Likert scales likely require such assumptions to be satisfied even to measure *ordinal* information about public opinion, if it is thought (as I have argued) that the underlying individual attitudes contain cardinal information. I have few definitive answers about what survey methodologists should conclude from this. However, when the purpose of public opinion information is to inform political representatives about the popular will, it should be remembered that the thing being measured has a *cardinal* structure. Insofar as public opinion surveys attempt to be informative in this way, survey methodologists should be concerned about how well the survey instruments actually take account of this underlying cardinal structure.

6.2.2 Decisions within deliberative institutions

Those systems of government recognized as democratic today are also typically deliberative. Most decisions in legislatures and other parts of government occur only after a substantive deliberative debate, followed by some voting procedure that provides closure to the decision. Such debates take place at multiple levels. The level that has probably received the most attention in the deliberative democratic literature is within the government, and especially the legislature.[20] However, other

[20]Although the legislature is often seen as the primary location of deliberation within government, deliberation that takes place within the bureaucracy of a government and between the bureaucracy and the legislature is also of importance. For instance, see Henry Richardson's (2002) recent discussion of the deliberative process both within the legislature, within the administrative bureaucracy, and between them.

levels of deliberation are also important to democratic theory. Deliberation among the people, either in formal "town-hall" type meetings or informally in cafes, bars or bowling alleys (to name a few), has been a part of the democratic traditions of many societies as well as democratic theory.[21] Equally important is a tradition of deliberation between the people and their representatives and public officials, again both formally and informally. Although I have already shown that such deliberative activity, a species of *direct communications*, cannot be counted as properly communicating public opinion to political representatives, it still plays an important function within modern democracies. These types of deliberation serve as the centerpiece of some versions of democracy. It would be good, therefore, to determine whether all of these types of deliberation are compatible with any representative decision process that I have proposed in this dissertation.

Legislatures are forced to combine aggregative and deliberative procedures when making policy decisions. It is not at all obvious how this can be done. Deliberative Democrats typically claim that decisions are ideally made through a rational consideration of all of the best arguments for the various courses of action. These arguments must be acceptable to all, leaving out consideration of purely private goods. Under this conception of political decision, it would seem that representatives should set aside the wishes of their constituents unless those wishes could be backed by arguments that are acceptable to all. Public opinion according to this view is not the *aggregative* conception I have presented here, but rather is a set of

[21]For instance, both Cohen and Fung's (2004) "Radical Democracy," which is a massively deliberative view of democracy, and Christiano's (1996) two step view of democracy that divides the labor between the citizens and the legislators, take the process by which the public deliberates about and forms opinions about policy issues to be central to democracy.

reasons for adoption of various policies that are presented by the public for evaluation by the deliberative body of the legislature. I do not think this respects public opinion as a set of considered judgments of the people. Rather, it demands that the judgments by the people be made *transparent* such that the reasons underlying them may enter into and be judged by the deliberative process within the legislature. And it is unclear how an appropriately deliberative process could function in any other way. It would seem that if representatives make use of opinion information itself, they must ignore their judgments about what the best arguments are or refuse to be moved by those arguments.

One reaction to the restriction of deliberation to reasoned argumentation is to expand the primary setting of deliberation to include all of society. This inverts the typical pattern of deliberation, such that many social decisions are made by citizens in local deliberative bodies. Under such a view, the role of elected officials may be reduced to introducing matters for consideration (for instance, Cohen and Fung (2004)) or guiding the deliberation Fishkin (2009). This is a radical reformation of the democratic systems that are common today, and on the face of it there is reason to doubt that such a system could adequately govern. However, it is not my purpose here to argue directly against such views, but just to show that it is not necessary to resort to such radical measures. There is a significant way in which political representatives can respect public opinion while at the same time respond to deliberative pressures within a legislative body.

I see public opinion as having two primary influences on rational deliberation within legislatures. First, political representatives have reasons to pursue arguments

in favor of alternatives endorsed by public opinion. The fact that a representative's constituents judge an alternative as being worth pursuing should provide reasons to a representative who respects their opinion; this opinion is derived from their judgments about what it is best to do, and if a representative does not currently understand why a certain alternative is worth pursuing, then she should attempt to understand this. An understanding of public opinion by a representative may arise through deliberation between the representative and her constituents. Opinion information itself should also sometimes be directly integrated into a representative's decision about how to vote after the deliberative process within the legislature has finished, especially when there is a strong opinion in favor of (or against) some alternative and the risks of adopting such a policy are not large. This arises from the balancing procedure I discuss in this dissertation, which treats public opinion as a set of deliberative judgments in themselves — a view that is bolstered by strong deliberative settings existing at the citizen-level.

In a healthy democracy, deliberation occurs both in the legislature as well as among the people. Deliberation allows the deliberators to better understand the rational basis for their policy opinions, and to revise them when appropriate. Without any sort of deliberation, it is far more difficult for individuals to understand possible opposing views, the arguments that support them, and consequently how to evaluate all the available positions and arguments. This is especially important if policy opinions should be directed at the common good, rather than simply each individual's (or group's) private good; it can be difficult to know the content of the common good if one does not understand the interests and viewpoints of others.

Of course, deliberation varies widely in quality, and this is especially the case in the distributed and informal deliberation that takes place within the public. If deliberative groups are internally homogenous enough, deliberation can succumb to various biases, including "groupthink," which tend to divert the deliberative process away from a careful analysis of all arguments on their merits.[22] Therefore, just as with deliberation among representatives, deliberation among citizens should include individuals with diverse political views and structured deliberative settings so that they deliberate in a way that promotes the rational consideration of all arguments. It cannot be expected that this deliberation will rise to the quality within well-functioning legislatures and bureaus, but a formal program of public deliberation may improve to some degree the quality of deliberation that takes place within the public.[23]

A third type of deliberation occurs between a representative and her constituents. This is importantly characterized as being asymmetric; the representative is in a position of authority, and consequently her views and arguments are provided an elevated position within the deliberation. Given the asymmetry, this type of deliberation might also be thought of as "education" of the public rather than pure deliberation. However, given that constituents typically *collectively* have the power to persuade a representative, I do not think that an education analogy is entirely appropriate. The degree to which representatives and others in leadership

[22]For instance, see Janis' (1982) classic analyses of groupthink biases in various group decision settings.

[23]Fishkin (2009), for instance, has conducted extensive research on the effects of deliberative sessions among average citizens of extremely varied backgrounds, showing some positive results on the extent to which citizens involved in the deliberations understand one another's concerns and arguments.

positions affect public opinion has been intensively studied by political scientists. For instance, see Berinsky (2009) for his analysis of the effects of leader attitudes on changes in citizen attitudes toward American wars. He finds the policy attitudes of leaders has little effect on those with little interest in policy issues and a very large effect on those who have great interest. Therefore, political leaders are often opinion leaders to those who look to them for political information. This echoes Zaller's (1992) more general claim that political elites significantly influence the policy opinions of the public about a number of topics. However, there is still significant room for public opinion to influence policy making, and a good deal of evidence that the public both acts as a *constraint* on the set of policies considered by representatives as well as a direct pressure on those policy decisions.[24] Leaders can have significant effect on public opinion, though needn't always, and are themselves susceptible to changes in public opinion that originate in the public itself. The causal link between a representative's and her constituents' opinions is bidirectional.

All three types of deliberation are compatible with, and likely required by, the *Duty of Political Responsiveness*. This is most easily seen when this duty is justified by its ability to permit self-governance. In order for the public to fully govern themselves, both their judgments and the reasons behind them must provide guidance to the deliberative and decision process. If political representatives failed to be responsive to either of these, they would be leaving their constituents out of some critical aspect of the decision process. As I argued in section 2.1.3, representatives

[24]See Jacobs et al. (1999). This is also similar to the view expressed by Page and Shapiro (1992) in their argument that public opinion is generally informed and rational.

cannot simply be responsive to the arguments or reasons of their constituents for their decisions to be suitably democratic. They must also respect the final judgments of their constituents. This is the balancing that I think is well represented in the *general will*, and which is behind the balancing decision procedure I defend in this dissertation. The existence of all three types of deliberation is therefore important for the instantiation of the *Duty of Political Responsiveness*.

There are many instances when knowing the reason for a constituent's opinion can help a representative determine how to use that opinion, and more concretely how to weigh it. Generally, a representative should weigh an opinion more when it seems backed by valid arguments, and when the point of disagreement between the constituency and the representative can be located in premises about which reasonable people could disagree. In such a case the representative can understand why her constituents made the judgments they did. In other cases, the representative may be able to locate clear instances of fallacious reasoning that are crucial to her constituent's judgments. Alternatively, the constituency may fall victim to collective irrationality, such as the *discursive dilemma*, in which sets of rational individual judgments may form an irrational set of judgments when aggregated (Kornhauser and Sager, 1993). These cases of explicit constituent irrationality should cause a representative to discount public opinion; in terms of the decision model, this means giving it much less weight. The most difficult cases are those in which the representative does not fully understand the reasoning behind her constituents' opinions, or thinks that they have unreasonable beliefs that have led them to some particular public opinion. However, in this case it is difficult for the representative to locate

a clear point of irrationality. Belief formation is a complex process, and it is often difficult to determine whether another's belief (or even our own) was arrived at through a reasonable process. Such opinions still require the representative's respect, and therefore should receive some weight, though likely less than in the first case I describe and more than in the second.

Deliberative theory plays a crucial role in understanding how political representatives should act to make democratic choices. I think that I have provided some insight into just what this role might be.

6.2.3 Implications for other areas of social choice theory

Before ending this chapter, I will point out some possible general implications that the method I have employed here has for social choice theory. In this dissertation, I have interpreted preferences and utility in a very specific way, but social choice theory provides very general results that do not require specific interpretations of preferences, individuals, utility, or its other elements. This interpretation neutrality is typically seen as a substantial benefit of the theory, because it makes its findings applicable within many different theoretical contexts. But I will argue that this generality comes at a cost. It often makes social choice theory unable to adequately answer certain types of questions in philosophy and the social sciences. For example, William Riker's (1982) now classic criticism of populism within political philosophy is based upon such a general analysis; he uses various results from social choice theory to show that voting behavior cannot possibly express some "popular

will" for the choice of candidates. Here I will cast doubt on the use of this sort of social choice theoretic analyses to come to such conclusions. In this dissertation I have presented a context-specific strategy for representing public opinion. Notice that this has a subject matter distinct from Riker's investigation; you might think that public opinion exists but not think that it is possible for any *election* result to reflect the preferences the public has for the candidates. The concept of public opinion that I have developed here is *passive*, such that it must be measured. However, it might be that voting behaviors instantiate the expression of a popular *will* of a different sort, a type that is *active* rather than passive. The fact that a government is backed by the active will of the people may be important for its legitimacy. I do not know whether this is the case, but its possibility provides reason to take seriously Riker's denial that such an active will exists.

I think that a general strategy related to what I have done in this dissertation has much promise in social choice theory, and in this section I will provide a more general framework for understanding it. In doing this I will discuss some other *context-specific* social choice models that have recently been proposed. Of particular interest is the work of Michael Regenwetter and his colleagues (2006), who have argued that the Arrovian problem of majority cycles in (ordinal) voting theory can be avoided by making statistical assumptions about electorates.[25] This circumvents

[25]A *majority cycle* in some triple occurs when there is no Condorcet winner in that triple. A Condorcet-winning alternative is one that wins a majority of votes in every pair-wise comparison with other alternatives. The majoritarian voting method finds a social ordering by first determining all pair-wise winners and then forming an ordering such that all alternatives are Condorcet winners against all alternatives ranked *below* them. A preference cycle, or a case of *intransitive* preferences, occurs when for some triple of alternatives x, y, z, all of the following pair-wise relations obtain: $x \succ y$, $y \succ z$, and $z \succ x$. Thus no Condorcet winner exists, and no social ordering can be found.

the sort of worries that are behind Riker's rejection of populism. This strategy is similar in kind to my own; by moving the social choice analysis into specific *theoretical contexts*, problems that are typically considered intractable in the standard (interpretation-neutral) social choice framework can often be solved. Therefore, our work points to an alternative view of social choice theory, and gives reason to be skeptical of many conclusions of context-general analyses.

Substantive theoretical assumptions are common in most of the sciences. These are used to set the basic theoretical background within which analyses can take place. I will argue that such assumptions are also often indispensable in social choice theory. It will first be helpful to distinguish between several types of assumptions, some shared by both context-free and context-specific social choice models, and others possessed only by context-specific models. All standard social choice models have both *structural* assumptions and *normative* assumptions.[26] The structural assumptions specify the appropriate *subject* of analysis, which for many theories means the basic rules that determine which preferences are admissible into a model and what counts as a situation covered by it; this is a standard use of the term "structural condition" in social choice theory (Fishburn, 1973).[27] In Arrovian social choice theory, the structural assumptions state the proper domain of a social welfare function and the output requirements of that function. Specifically, the theory re-

[26]Although the *identity* of the model is defined *extensionally*, by the set of *functions* that the model takes under consideration, these models are typically characterized intensionally, by the set of *assumptions* that constrain the set of functions under consideration. Therefore, even though the assumptions that I will discuss do not themselves *define* social choice models, they are required for any actual description and application of those models.

[27]Note that I use the word 'assumption' instead of 'condition', which is more common in social choice theory, though I use them equivalently.

quires that there be a society with at least two individuals who all have weak orders over three or more alternatives. Additionally, a social welfare function (the output) must be a weak order.[28] An alternative social choice model might only require that a "social *choice* function" be found, such that the output is just a set of the winning alternative(s).[29] But this is a different social choice method.

It is useful to classify other assumptions as *normative* because they posit certain requirements on an acceptable social decision, but are not basic to the subject matter of analysis. For instance, Arrow's *Independence of Irrelevant Alternatives* requires that the social ranking of any two alternatives with respect to one another should not be affected by the ranking of some third alternative. Those who support the inclusion of this condition typically take a decision that is affected by such "irrelevant" third alternatives to be perverse in some way.[30] Other standard Arrovian normative assumptions include some version of a *Pareto Principle* and *Non-Dictatorship*. Notice that it is possible to conceptualize these conditions as

[28] x is weakly preferred '$\succeq$' to y iff x is preferred at least as much as y. This is similar to a '$\geq$' relation for numerical values.

[29] See Sen and Pattanaik (1969). In this paper, I will primarily use the term "social choice function" in its more general sense, to refer to any function satisfying the requirements of social choice theory, including "social welfare" and "social choice" functions.

[30] Arrow's Impossibility Theorem is one of the central findings of social choice theory. It shows that no aggregation procedure (or voting method) guarantees a coherent *social* ordering (or ranking) of alternatives if it is assumed that individuals may have any possible individual orderings of the alternatives and that the aggregation procedure must satisfy a set of reasonable conditions on democratic choice. These conditions are (where x and y are two alternatives):

Weak Pareto Efficiency If all individuals rank x above y, then the social ordering must rank x above y.

Independence of Irrelevant Alternatives The social ranking of any two alternatives x and y should only depend upon the *pair-wise* ranking of those alternatives. The relative ranking of other alternatives is *irrelevant* to the social ranking of x and y.

Non-Dictatorship There is no individual i, such that no matter how other individuals rank x and y, the ranking of x and y by i becomes the social ranking.

structural, but doing so strains the limits of what would generally be considered the basic structure of the theory. The reason why *Non-Dictatorship* is an important assumption in the theory is not because it is integral to the basic problem of social decision, but rather because any social decision rule that did not satisfy it would fail to be democratic. Normative assumptions specify how decision rules should be selected, given some basic social choice theoretic framework. This permits a normative debate about which decision rules are properly democratic once the basic theory parameters are specified through a set of structural assumptions.

A common assumption in social choice theory that nicely illustrates the distinction between structural and normative assumptions is the restriction of the choice *domain*.[31] This is a restriction on the preference profiles that are admissible into the social welfare function. It is more naturally a *structural* assumption, because it restricts the possible inputs to the social decision, which changes the social choice problem itself. One might instead think of such domain restrictions as primarily normative in nature, but it is doubtful that such a normative assumption could be justified. Generally, it is thought that individuals should be able to have any preferences whatsoever. Therefore, a normatively justified domain restriction would require a rationale that makes use of particular normative features of a theoretical context, such that it would be permissible within the context to restrict the preferences of some individuals because they do not conform to the domain restriction. Because such normative justifications are not obvious within most contexts, it seems

[31]See Black (1958) and Sen and Pattanaik (1969) for classic presentations of domain restrictions. Also see Le Breton and Weymark (2010) for a summary of such restrictions compatible with economic goods, and Gaertner (2006) for restrictions in collective choice and voting theory.

unlikely that domain restrictions could be justified through context-free or context-specific normative grounds. The theoretical context under consideration sometimes gives reason for thinking that certain preference profiles will not (or are unlikely to) arise, but does not provide reasons why some profiles should be ignored even if they do arise. Domain restrictions then should be conceived of as context-specific structural assumptions.

Though most work in social choice theory is relevant to a number of different contexts, that does not eliminate a role for numerous competing social choice models, some of which might perform better in some contexts than others. However, a key feature of context-free social choice models is their independence from any particular *interpretation* of the entities in question. Context-*specific* models, on the other hand, require a particular interpretation of the model elements to justify the assumptions within the theory. Allowing a social choice model to be restricted to some particular context in this way opens up a possibility of reasonable *context-specific* structural and normative assumptions in addition to, or in place of, context-free assumptions.

Context-free social choice models tend to produce more skeptical results than context-specific ones due to the relative austerity of context-free assumptions. For instance, context-*free* social choice models have generally supported political theories in political science that are skeptical of popular democratic rule. Perhaps most famously, William Riker (1982) argues that several results from social choice theory show that the populist theory of democracy is unworkable, because there can be no coherent "popular will" that populist theories require public decisions to track. This strong result arises from several related findings of context-free social choice

theory, showing that no voting rule could possibly satisfy the requirements of democratic choice (Arrow's Theorem), that every election is subject to strategic voting (Gibbard – Satterthwaite Theorem), and that control over the agenda of an election has significant effects on its result. The plausibility of Riker's argument rests on the applicability of the context-free assumptions of social choice theory, and *only those assumptions*, to the particular theoretical context of democratic decision making within large societies.

Given that widespread applicability is one of the strengths of such context-free theories, it seems reasonable that context-free social choice theory should be able to support substantive findings like Riker's. However, the austerity of context-free assumptions in standard social choice theory makes skeptical findings more likely. Therefore, the findings Riker cites represent the *worst*-case scenario for democracy, not a likely one. But moving beyond austere context-free assumptions requires a different type of argument — one that is reliant upon particular features of the phenomena of interest in addition to the general principles upon which context-free social choice models are based. In this dissertation I have provided a social choice theoretic framework that permits interpersonal comparisons of cardinal utility scales when those scales represent *attitudes*. The assumptions I made about the underlying attitudes those utility scales represent were necessary for those comparisons. I will also examine recent work by Regenwetter et al. (2006) on the subject of how social decisions are possible in large-scale elections. In both of these contexts (large-scale elections and cardinal public opinion measurement) context-free social choice models come to incorrect skeptical conclusions, whereas context-specific models do not.

Cyclic social preferences are possible when using the majoritarian voting method (aka. The Condorcet Method), which is the only voting method satisfying Arrow's (normative) conditions on democratic choice. In the standard Arrovian social choice model, such cycles are highly likely as the number of alternatives and/or individuals becomes large, approaching 1 as they approach infinity (Riker, 1982, pg. 122). Regenwetter et al.'s model relies upon statistical assumptions related to restrictions on the *domain* of the social aggregation function, and they argue that actual voting patterns (voting *profiles*) from real-world electorates are structured in ways that should be characterized within the social choice model. When these assumptions about the structure of electoral preferences are made, majority cycles are extremely unlikely, even over large sets of alternatives and in large electorates. Importantly, the assumptions about the voting patterns of the electorate are *empirical* in nature, and are only plausible given the truth of certain empirical claims.

The context-specific assumptions of Regenwetter et al. (2006) are empirical in nature, but this need not be the case within all contexts or for all assumptions. In fact, a simpler domain restriction than what Regenwetter et al. consider within the context of *scientific theory selection* has recently been suggested that may overcome any Arrovian problems based purely upon *conceptual* relationships between different scientific merits.[32] Arrow's theorem applies in nearly any case of multidimensional choice, even in the selection of best scientific theories using various scientific merits (Okasha, 2011). For instance, if three scientific merits are used to rank alternative theories (logical strength, generality, and fit to the data), then Arrow's theorem

[32]See Morreau (2013).

shows that there is no adequate method to find an aggregate ranking of theories. However, if relationships can be found between the scientific merits, then the *domain* of such a theory choice method might be restricted, making a majority decision rule possible. This constitutes a *conceptual* structural assumption that is included in the context-specific model for theory selection. For example, the logical strength and fit to the data of a scientific theory are often construed such that they form inverse rankings of the alternative theories (Morreau, 2013). Therefore, if the theories x, y, z are ranked $x \succ y \succ z$ by *logical strength*, then those theories must be ranked $z \succ y \succ x$ in terms of how well they *fit* the data. This places increased structure on the domain of the aggregation function, which makes a majority preference cycle less likely. In fact, if it is assumed that there is only one additional scientific merit besides logical strength and fit, then this restriction on the domain satisfies Sen's *Value Restriction* (Sen and Pattanaik, 1969), which is a sufficient condition of guaranteeing that a majority cycle is avoided. *Value restriction* is also a structural assumption important to Regenwetter et al.'s (2006) argument that large-scale elections are not typically susceptible to majority cycles.

Regenwetter et al. (2006) claim that real-world electorates generally conform to a statistical variant of Sen's *Value Restriction* condition, and thus it is extremely unlikely that any given election will end in a majority cycle. Sen's solution to the problem of majority cycles is to drop Arrow's requirement of *Universal Domain* on the set of possible preference *profiles*. Sen's *Value Restriction* requires that, within every triple of alternatives, there is consensus on what rank *some* alternative does *not* have. Therefore, within each triple, there is some alternative that satisfies one

of the following conditions: never ranked first, never ranked middle, or never ranked last. This condition is sufficient for a profile of weak orders to avoid a majority cycle, and is necessary and sufficient for a profile of *strict* orders to guarantee the avoidance of a majority cycle; importantly, these guarantees apply to the profile regardless of whether some individuals are duplicated or eliminated. I will return to this last point, because it is important for distinguishing Sen's *Value Restriction* from Regewetter et al.'s assumptions. Regenwetter et al. show that *Value Restriction* is rarely satisfied for real-world electorates, and therefore is not a suitable structural assumption for voting contexts. However, they revise Sen's *Value Restriction* assumption, replacing it with a *statistical* assumption over possible preference profiles. This revised statistical assumption is supported by the available empirical evidence, though it only assures that majority cycles are *highly unlikely* rather than impossible.

According to Regenwetter et al. (2006), two separate arguments show the unlikelihood of majority preference cycles in real elections. First they note that the worst-case scenario for determining the composition of preference profiles is a uniform distribution, or what is called an "impartial culture" over possible individual preference orderings. This is precisely what Riker assumes when he makes his well-known calculations showing that majority cycles are extremely likely in all except very small electorates (Riker, 1982, pg. 122). Regenwetter et al. show that if it is simply assumed that some preference orderings are more likely than others, without even claiming anything specific about what that new probability distribution is, then majority cycles are not as likely as Riker claims. The plausibility of this

denial of *impartial culture* can be intuitively seen in typical national elections. For example, in the 2008 American presidential election (including both primary and general elections), it would have been more likely for someone to rank the candidates Obama $\succ$ Clinton $\succ$ McCain than Obama $\succ$ McCain $\succ$ Clinton. This gets to the core of the context-specific strategy; a denial of impartial culture in the case of elections requires that there are certain regularities in individual preferences over the candidates. It is obvious that this is the case in most elections, but such an assumption would not be defensible were the social choice model not restricted to this particular theoretical context.

Next Regenwetter et al. argue that *even if* the culture is impartial, something that is likely not true, a statistical assumption about the distribution over possible preference *profiles* makes majority preference cycles very unlikely. They present empirical evidence that their statistical variant of *Value Restriction* is satisfied for many real-world electorates, and this allows them to argue that real-world electorates have only small probabilities of producing majority cycles. They call this assumption *Net Value Restriction*. This restriction is distinct from Sen's *Value Restriction* in two primary ways. First, and most importantly, it is a restriction on *profiles* rather than combinations of orders that are instantiated within particular profiles. Sen's *Value Restriction*, therefore, can fail to be satisfied by an infinite number of possible profiles that nonetheless do not lead to majority cycles. For example, the following profile does not satisfy *Value Restriction*, but does not lead to a majority cycle:

$x \succ y \succ z$

$y \succ z \succ x$

$z \succ x \succ y$

$z \succ x \succ y$

$z \succ x \succ y$

The order $z \succ x \succ y$ receives a majority and thus there is no majority cycle. This fails *Value Restriction*, because there is some subset of the profile (the first three orders) in which no alternative is never ranked first, never ranked middle, or never ranked last. Of course, the profile in which $z \succ x \succ y$ obtains a majority is a somewhat trivial case of avoiding a majority cycle, but makes it easy to see why assumptions about which individual *profiles* are likely (or unlikely) provide greater analytic power than those that do not. Regenwetter et al.'s *Net Value Restriction* assumes such a distribution over the space of profiles themselves, and consequently they are able to make claims about the probability of various profiles that would not be possible if only the sets of possible orders (rather than profiles themselves) were subject to restriction.

The second way in which Regenwetter et al.'s *Net Value Restriction* differs from Sen's *Value Restriction* is in its statistical formulation. Whereas *Value Restriction* provides a sufficient condition on there being a social ordering, such that a social ordering will *certainly* be possible if the condition is satisfied, Regenwetter et al.'s condition only claims that a majority cycle has some low probability of aris-

ing from a profile exhibiting *Net Value Restriction*. Furthermore, their empirical evidence only permits them to say that there is some high probability of *Net Value Restriction* being satisfied in real-world electorates. The relative weakness of this claim, for instance as compared to those I describe above from theory selection, is a result of the *empirical* rather than *conceptual* context-specific structural assumptions. Whereas the merits of scientific theories may possess certain relations that allow for non-statistical assumptions such as *Value Restriction*, empirical phenomena typically do not. Therefore, any social choice model of elections that makes use of what *is* known about the electorate will typically only be able to come to conclusions that have certain (high) probabilities of being true. This is a deficit of a social choice model only in the way that similar probabilistic findings are deficits of models in biology, psychology, or economics. And it dramatically changes the basic method of social choice theory, and results in a scientific extension of the purely formal methods found in context-free social choice models. This also gives an indication of why *Value Restriction* may be successful in the case of theory selection but unsuccessful for elections. Whereas the phenomenon of theory selection may be well described by purely abstract *a priori* assumptions, this is not the case for voting behaviors. The patterns of preferences typically held by specific electorates can only adequately be known by reference to empirical data, which dooms any use of standard domain restrictions within the theoretical context of elections.

I have discussed how the use of context-specific social choice models can more adequately address a number of classic problems in social choice theory and related fields. The upshot of this discussion is that social choice theorists often make *fewer*

assumptions about the underlying phenomena they are modeling than they are entitled to; this sometimes leads to overly skeptical conclusions. I have presented two primary examples of context-specific social choice models, one from the political science literature and another that is novel. They both allow for the avoidance of skeptical findings within particular theoretical contexts. First, I presented recent work by Regenwetter et al. (2006) that shows how the likelihood of a majority social preference cycle can be minimized if a number of statistical assumptions are made. They argue that these assumptions are likely satisfied by most real-world electorates. I then presented a similar strategy to resolve problems in utility theory surrounding the interpersonal comparison of cardinal utilities, at least when those utilities are used to represent *attitudes*.

The example of interpersonal comparisons of utility that I have given in this dissertation is particularly useful for understanding what benefits can be obtained from context-specific models. These benefits lie primarily in shifting the focus of social-choice theoretic questions to a more *concrete* theoretical context; this then fills in the abstract structure of findings in context-free social choice theory, showing what follows from its general principles given particular *interpretations*. The formal problem of making interpersonal comparisons of utility is, in many ways, intractable without the move to a particular context. Context-free social choice and utility theories provide the formal constraints on the debate, but not enough to arrive at firm conclusions. For instance, the types of interpersonal comparisons and how they relate to the measureability of individual scales are both necessary antecedents to any context-specific model of those comparisons, but context-free utility theory

says little about whether interpersonal comparisons are possible within any given situation. Therefore, many substantive debates must take place within particular contexts.

I have suggested two main types of assumptions that can be made within theoretical contexts: structural and normative. These are also the two main areas of debate within theoretical contexts. Referencing my own example of interpersonal comparisons of cardinal attitudes, one might disagree with the assumption that stable attitudes are connected in the manner I presented to pair-wise choices between public policies; perhaps responses are almost entirely the result of factors *external* to a person's psychology. This would render unfounded the individual utility scaling technique I employ. One might also object to the structural assumptions that permit interpretations of individual utility scales and subsequent interpersonal comparisons of attitudes. It is plausible that one might, for instance, deny that individual dispositions can be characterized by variants of the normal distribution. Alternately, one might deny that it makes sense to make any dispositional claims about individual utility assignments within the context of attitude measurement. The resolution of *that* question would likely require an investigation into how individual attitudes arise from underlying psychological processes, which would require a predominantly empirical investigation. It makes little sense for any of these debates to take place outside of a particular theoretical context.

Normative assumptions invite debate over how utility and social choice theoretic information should be used, rather than about the basic structure of the phenomena being represented. In the case I consider in this dissertation, where

public opinion is used to inform the decisions of political officials, the commitment to one set of normative assumptions requires further commitments to particular views about political representation and the roles of fairness and responsiveness in democratic decisions. I pursued these normative questions in chapters 2 and 3. I have claimed that if public opinion is to be used by political representatives, for instance legislators, then two normative assumptions should likely guide its use: *Overall Fairness* and *Special Interests*. Notice that an argument for this must take place within the current constraints of political philosophy, which would in part specify the content of a theoretical context. In other words, this relies upon a *context-specific* model.

Chapter 7: Conclusion

I have argued in this dissertation that political representatives have duties to be responsive to public opinion in their public policy decisions. How exactly they do this is often complex, and can vary depending upon the exact situation, but the guidance offered by this duty is more obvious in some cases than others. Consider the example of the Affordable Care Act (ACA) that I began with in the Introduction. For many political representatives, it will be difficult to determine whether a vote for or against the act constitutes a violation of that representative's duty to respect public opinion. This just reflects the complexity of the cases involved. Policy decisions are often difficult for both the public and representatives, with the motives of the various actors involved obscured and important facts sometimes impossible to sufficiently verify. But I think there are some clear cut examples that show how the *Duty of Political Responsiveness* might be used to guide and evaluate the acts of representatives in instances of real-world policy making. .

Consider a Democratic U.S. Senator A from a conservative state, who is faced with a decision of whether to vote for the ACA. Imagine that there has been extensive opinion polling done in her state, and she concludes that a large majority is strongly opposed to the passage of the ACA. Although in this historical case, the

public opinion polling information likely fails to meet the standards that I have set in this dissertation, for the purposes of this discussion it might be assumed that the polling information is either of extraordinary quality or that the opinions that are measured are unambiguous. For instance, there may be so few people who support the ACA that their cardinal opinions could not possibly overcome the opinions of those opposed in any defensible aggregation function (like *normalize and average*).

In the above case, it is clear that certain considerations should be excluded from A's deliberations about her decision. First, she should not include considerations of pure party loyalty unless she thinks that loyalty significantly furthers the common good. Similarly, she must also exclude considerations of personal advancement within the Senate. The exclusion of these considerations will likely be recommended by many theories of political representation, but some others are particular to my own view. For instance, A should generally not take electoral outcomes as conveying a "mandate" to govern according to her electoral platform. As I have argued, elections results are not capable of conveying this support for a representative's electoral platform; public opinion survey information provides better information than election results. A's view about what is the best policy to pursue must be weighed in her decision according to her estimation of the extent of negative or positive results of pursuing that policy, along with how sure she is of that estimation; this should reduce the importance of purely ideological policy commitments that the representative has little independent reason to hold.

The Duty of Political Responsiveness sometimes allows for public opinion to be largely ignored when the representative has good reason to think that acting accord-

ing to public opinion could produce extremely negative consequences for the public. Consider Representative B, who is deciding whether to support the Emergency Economic Stabilization Act of 2008. This bill provided financial aid to distressed banks during the American financial crisis of that year. B is a Democrat whose constituents are overwhelmingly against the law; she receives numerous phone calls and emails from constituents, many of who say that it is wrong to let banks off "scot-free" while making the middle class foot the bill. More formal public opinion surveys convey a similar message. Most of her constituents do not seem to think that more serious harm will befall the economy if banks are left without any assistance. However, she has had an opportunity to sit in a number of meetings with experts in economics and banking, and those experts were in agreement that if the law were not passed then there would likely be serious negative repercussions on the American economy. That convinces her that she must vote for the bill. Her choice is consistent with *The Duty of Political Responsiveness*, because her decision takes into account public opinion though puts less weight on it. She only rejects the choice advocated by the opinion of her constituents upon realizing the significant risk of not approving the bill and the degree to which it seems that the public does not significantly take into account this risk when its members formulate their opinions.

Another case will show the importance of the cardinal conception of public opinion. Governor C, who is a Republican, is determining whether she should expand her state's Medicaid system to include those individuals just above the poverty line. She is inclined to deny the expansion of this program, because she is suspicious of more Federal involvement in her state's finances and because it would

improve her image among party leaders and donors. There is also opinion polling information showing a slight majority of her constituents are also against expanding Medicaid. However, this polling information, along with her experience speaking to various groups of citizens, leads her to believe that a significant minority in her state *strongly* favor the expansion of Medicaid; this minority includes both those who are directly affected by the potential policy change as well as those who are aware of and sympathetic to those directly affected. This is evidence that the strength of preference of those who are against the expansion of Medicaid is far weaker than the strength of preference for those who favor it, even though there are more people who are against it than who favor it. The *Duty of Political Responsiveness* may require C to consider expanding Medicaid even though most of her constituents do not want the policy to be implemented and she is also inclined against it. This relies upon the relative weakness of both her own disfavor towards it as well the weak preferences of those in her constituency against it. If however she, for instance, has good reason to think that expanding Medicaid would be harmful to her state's budget then this might provide more strength to her belief that it would harm her constituents. This might then be enough to resist the stronger opinions of the minority who favors the policy.

Finally, it is important to see how the duty I argue exists will sometimes require decisions that run contrary to the institutional incentives that provide representatives with pragmatic reasons to act in particular ways. State Senator D is a Republican from a district with two distinct populations. 60% of her constituents live in a more urban area and are relatively young, poor, and liberal. The other 40%

of her constituents live in a suburban area and are relatively old, rich, and conservative. However, those in the suburban part of her district tend to vote at much higher rates than those who live in the urban part of her district; this is compounded by the fact that all of her state elections occur in non-Presidential election years, when the turnout is especially low for those living in her urban districts. Because D is concerned about a primary challenge, and relatively unconcerned about a challenge in a general election, she tends to support especially conservative policies, even though a majority of her constituents typically support more liberal policies. In this case, D is acting to maximize her chances of being reelected, and the particular institutional incentives in place make it (pragmatically) rational for her to support conservative policies that most of her constituents do not. This fails to act according to the *Duty of Political Responsiveness*, which is especially easy to see given the generality of her tendency to ignore public opinion.

The wrongness of D's decision procedure makes clearer the relationship between what I have done in this dissertation and the work in political theory about institutional design. Much of the concern in political theory is (rightly, I think) concerned with designing institutions that make desirable actions rational for the agents within those institutions. This dissertation has been concerned with the rightness and wrongness of decisions by political representatives without regard to other institutional pressures those representatives face. I find it uncontroversial that some of our duties will sometimes (and perhaps often) require actions to the contrary of the various incentives that are relevant to our decisions. This occurs with political representatives as well. However, the existence of the duty, especially

one so central to political representation, gives some guidance to how institutions should be designed. Better political institutions will tend to incentivize the sort of behaviors that are consistent with the duties possessed by political officials. Just as political institutions should be designed to limit corruption by public officials, they should also be designed to give representatives incentives to properly represent their constituents.

In this dissertation I have argued that political representatives have duties to respect public opinion. Showing what constitutes "respect" and "public opinion" required formal models of both political decision making and public opinion. These allowed me to show that there is nothing deeply wrong with the concept of public opinion that renders it impossible to either measure or include in public opinion in public decision making. Additionally, I showed how public opinion can be modeled as a cardinal utility scale, such that it is possible to weigh it against other factors in decision making; most important of these factors is a representative's estimation of the likely effects of the policy, which is often modeled using cost-benefit analysis or other cardinal metrics of welfare. This I think provides a framework for evaluating the actions of political representatives.

Bibliography

Achen, C. H. (1975). Mass political attitudes and the survey responses. *American Political Science Review 69*, 1218.

Achen, C. H. (1978). Measuring representation. *American Journal of Political Science 22*(3), 475–510.

Althaus, S. L. (2003). *Collective Preferences in Democratic Politics*. Cambridge University Press.

Anscombe, F. J. and R. J. Aumann (1963, 03). A definition of subjective probability. *Ann. Math. Statist. 34*(1), 199–205.

Arrow, K. (1977). Extended sympathy and the possibility of social choice. *The American Economic Review 67*, 219–225.

Arrow, K. J. (1963). *Social Choice and Individual Values* (2 ed.). John Wiley & Sons.

Berinsky, A. (2009). *In Time of War: Understanding American Public Opinion from World War II to Iraq*. University of Chicago Press.

Black, D. (1958). *The Theory of Committees and Elections*. Cambridge University Press.

Bossert, W. and J. A. Weymark (2004). Utility in social chioce. In S. Barbera, P. J. Hammond, and C. Seidl (Eds.), *Handbook of Utility Theory: Extensions*, Volume 2, Chapter 20, pp. 1099–1177. Kluwer Academic Publishers.

Brams, S. and A. Taylor (1996). *Fair Division: From Cake-Cutting to Dispute Resolution*. Cambridge University Press.

Brennan, G. and A. Hamlin (2000). *Democratic Devices and Desires*. Cambridge University Press.

Brennan, G. and L. E. Lomasky (1997). *Democracy and decision: The Pure Theory of Electoral Preference*. Cambridge University Press.

Broome, J. (1999). Utility. In *Ethics out of Economics*. Cambridge University Press.

Broome, J. (2013). *Rationality Through Reasoning*. Wiley.

Caplan, B. (2008). *The Myth of the Rational Voter: Why Democracies Choose Bad Policies (New Edition)*. Princeton University Press.

Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-knowledge*. Oxford University Press.

Christiano, T. (1996). *The Rule of Many*. Westview Press.

CNN (2010, May). May 22, 2010 cnn opinion research poll. http://i2.cdn.turner.com/cnn/2010/images/03/22/rel5a.pdf.

Cohen, J. (1989). Deliberation and democratic legitimacy. In A. Hamlin and P. Pettit (Eds.), *The Good Polity*. Blackwell.

Cohen, J. (1999). Reflections on habermas on democracy. *Ratio Juris 12*, 385–416.

Cohen, J. and A. Fung (2004). Radical democracy. *Swiss Political Science Review 10*, 23–34.

Cohen, J. and C. Sabel (1997). Directly-deliberative polyarchy. *European Law Journal 3*, 313–342.

Converse, P. E. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), *Ideology and Discontent*. Free Press.

Converse, P. E. (2006). Democratic theory and electoral reality. *Critical Review: A Journal of Politics and Society 18*, 297–329.

Converse, P. E. and G. Markus (1979). Plus ça change. . . the new cps panel study. *American Political Science Review 73*, 32–49.

Dahl, R. A. (1991). *Democracy And Its Critics*. Yale University Press.

DeCanio, S. (2006). Mass opinion and american political development. *Critical Review: A Journal of Politics and Society 18*, 143–155.

Dietrich, F. and C. List (2013). A reason-based theory of rational choice. *Nous 47*, 104–134.

Downs, A. (1957). *An Economic Theory of Democracy*. Harper & Row.

Drasgow, F., O. S. Chernyshenko, and S. Stark (2010). 75 years after likert: Thurstone was right! *Industrial and Organizational Psychology 3*, 465–476.

Eagly, A. H. and S. Chaiken (1993). *The Psychology of Attitudes*. Harcourt Brace Jovanovich College Publishers.

Esmer, Y. and T. Pettersson (2007). The effects of religion and religiosity on voting behavior. In R. J. Dalton and H.-D. Klingemann (Eds.), *The Oxford Handbook of Political Behavior*, Chapter 25, pp. 481–503. Oxford University Press.

Falmagne, J. C. (1985). *Elements of Psychophysical Theory*. Oxford University Press.

Fishburn, P. (1973). *The Theory of Social Choice*. Princeton University Press.

Fishburn, P. C. (1991). Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty 4*(2), 113–134.

Fishkin, J. (2009). *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford University Press.

Frey, William, H. (2013, May). Minority turnout determined the 2012 election. http://www.brookings.edu/research/papers/2013/05/10-election-2012-minority-voter-turnout-frey.

Friedman, J. (2006). Democratic competence in normative and positive theory: Neglected implications of "the nature of belief systems in mass publics". *Critical Review: A Journal of Politics and Society 18*, 1–43.

Gaertner, W. (2006). *Domain Conditions in Social Choice Theory*. Cambridge University Press.

Gallup, G. and S. F. Rae (1940). *The Pulse of Democracy: The Public-Opinion Poll and How It Works*. Simon and Schuster: New York.

Gibbard, A. (1973). Manipulation of voting schemes: a general result. *Econometrica 41*, 587–601.

Goldman, A. I. (1995). Simulation and interpersonal utility. *Ethics 105*, 709–726.

Goodin, R. E. (1995). Laundering preferences. In R. E. Goodin (Ed.), *Utilitarianism as a Public Philosophy*, pp. 75–101. Cambridge University Press.

Habermas, J. (1992). *Between Facts and Norms*. MIT Press.

Hamilton, A., J. Madison, and J. Jay (1788/2004). *The Federalist Papers*. Simon and Schuster: New York.

Harsanyi, J. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy 63*, 309–321.

Harsanyi, J. (1982). Morality and the theory of rational behaviour. In A. Sen and B. Williams (Eds.), *Utilitarianism and Beyond*, Chapter 2, pp. 39–62. University of Cambridge Press.

Hepler, J. and D. Albarracin (2013). Attitudes without objects: Evidence for a dispositional attitude, its measurement, and its consequences. *Journal of Personality and Social Psychology 104*, 1060–1076.

Horty, J. (2012). *Reasons as Defaults*. Oxford University Press.

Jacobs, L. R., C. J. Glynn, S. Herbst, G. J. O'Keefe, and R. Y. Shapiro (1999). Public opinion and policymaking. In *Public Opinion*, Chapter 9, pp. 299–340. Westview Press.

Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education 38*(12), 1217–1218.

Janis, I. (1982). *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Houghton Mifflin.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.

Knutsen, O. (2007). The decline of the social class? In R. J. Dalton and H.-D. Klingemann (Eds.), *The Oxford Handbook of Political Behavior*, Chapter 24, pp. 457–480. Oxford University Press.

Kornhauser, L. A. and L. G. Sager (1993). The one and the many: Adjudication in collegial courts. *California Law Review 81*, 1–59.

Krosnick, J. A. and R. P. Abelson (1994). The case for measuring attitude strength in surveys. In J. M. Tanur (Ed.), *Questions about Questions*, Chapter 10, pp. 177–203. Russell Sage Foundation.

Kymlicka, W. (2002). *Contemporary Political Philosophy: An Introduction* (2 ed.). Oxford University Press.

Lackey, J. (2008). What should we do when we disagree? In T. S. Gendler and J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, Volume 3, pp. 274–293. Oup.

Le Breton, M. and J. A. Weymark (2010). Arrovian social choice theory on economic domains. In K. J. Arrow, A. Sen, and K. Suzumura (Eds.), *Handbook of Social Choice and Welfare*, Volume 2. Elsevier.

Lewis-Beck, M. S. and M. Stegmaier (2007). Economic models of voting. In R. J. Dalton and H.-D. Klingemann (Eds.), *The Oxford Handbook of Political Behavior*, Chapter 27, pp. 518–537. Oxford University Press.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology 22*, 1–55.

Lippmann, W. (1922). *Public Opinion*. Harcourt, Brace.

Luce, R. D. (1959). *Individual Choice Behavior*. John Wiley & Sons.

Lyon, A. and E. Pacuit (2013). The wisdom of crowds: Methods of human judgement aggregation. In P. Michelucci (Ed.), *Handbook of Human Computation*, pp. 599 – 614.

Mayhew, D. (1974). *Congress: The Electoral Connection*. Yale University Press.

McDonald, M. P. (2012, December). *Presidential Voter Turnout Rates, 1980-2012*. United States Elections Project.

Mill, J. S. (1859/1978). *On Liberty*. Hackett.

Morreau, M. (2013). Mr. fit, mr. simplicity and mr. scope: From social choice to theory choice. *Erkenntnis 79*, 1–16.

Morris, C. W. (2002). *An Essay on the Modern State*. Cambridge University Press.

Nosek, B. A., M. Banaji, and A. G. Greenwald (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice 6*(1), 101.

Okasha, S. (2011). Theory choice and social choice: Kuhn versus arrow. *Mind 120*(477), 83–115.

Page, B. I. and R. Y. Shapiro (1992). *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. The University of Chicago Press.

Peffley, M. and R. Rohrschneider (2007). Elite beliefs and the theory of democratic elitism. In R. J. Dalton and H.-D. Klingemann (Eds.), *Oxford Handbook of Political Behavior*, Chapter 4, pp. 65–79. Oxford University Press.

Peterson, M. (2008). *Non-Bayesian Decision Theory: Beliefs and Desires as Reasons for Action*. Springer.

Pettit, P. (1997). *Republicanism*. Oxford University Press.

Petty, R. E., R. H. Fazio, and P. Briñol (2012). *Attitudes: Insights from the New Implicit Measures*. Psychology Press.

Pitkin, H. F. (1967). *The Concept of Representation*. University of California Press.

Pivato, M. (2013). Formal utilitarianism and range voting. *Mathematical Social Sciences 67*, 5056.

Przeworski, A. (1999). Minimalist conception of democracy: a defense. In I. Shapiro and C. Hacker-Cordon (Eds.), *Democracy's Value*, pp. 23–55. Cambridge University Press.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.

Raz, J. (1999). *Practical Reason and Norms* (2 ed.). Oxford University Press.

Regenwetter, M., B. Grofman, A. A. J. Marley, and I. M. Tsetlin (2006). *Behavioral Social Choice*. Cambridge University Press.

Richardson, H. S. (2002). *Democratic Autonomy: Public Reasoning about the Ends of Policy*. Oxford University Press.

Riker, W. (1982). *Liberalism Against Populism*. W. H. Freeman and Co., San Francisco.

Robbins, L. (1952). *An Essay on the Nature and Significance of Economic Science*. Macmillian and Co., Limited.

Roberts, J. S., J. E. Laughlin, and D. H. Wedell (1999). Validity issues in the likert and thurstone approaches to attitude measurement. *Educational and Psychological Measurement 59*, 211–233.

Rousseau, J.-J. (1762/1987). *On the Social Contract*. Hackett.

Saggar, S. (2007). Race and political behavior. In R. J. Dalton and H.-D. Klingemann (Eds.), *The Oxford Handbook of Political Behavior*, Chapter 26, pp. 504–517. Oxford University Press.

Satterthwaite, M. (1975). Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory 10*(2), 198–217.

Schuman, H. and S. Presser (1996). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Quantitative Studies in Social Relations. SAGE Publications.

Schumpeter, J. A. (1950). *Capitalism, Socialism, and Democracy* (3rd ed.). Harper & Brothers.

Sen, A. K. (1970). Interpersonal aggregation and partial comparability. *Econometrica 38*, 393–409.

Sen, A. K. (1979). Interpersonal comparisons of welfare. In M. Boskin (Ed.), *Economics and Human Welfare*. Academic Press, Inc.

Sen, A. K. (1992). *Inequality Re-examined*. Oxford: Clarendon Press.

Sen, A. K. (1995). Rationality and social choice. *The American Economic Review 85*(1), 1–24.

Sen, A. K. and P. K. Pattanaik (1969). Necessary and sufficient conditions for rational choice under majority decision. *Journal of Economic Theory 1*, 178–202.

Simmons, A. (2001). *Justification and Legitimacy: Essays on Rights and Obligations*. Cambridge University Press.

Sreenivasan, G. (2000). What is the general will? *Philosophical Review 109*(4), 545–581.

Stasavage, D. (2004). Open-door or closed-door? transparency in domestic and international bargaining. *International Organization 58*, 667–703.

Stevens, S. S. (1946). On the theory of scales of measurment. *Science 103*, 677–680.

Suppes, P., D. H. Krantz, D. R. Luce, and A. Tversky (1989). *Foundations of Measurement: Geometrical, Threshold, and Probabilistic Representations*. Academic Press, Inc.

Surowiecki, J. (2005). *The Wisdom of Crowds*. Knopf Doubleday Publishing Group.

Thurstone, L. L. (1959). *The Measurement of Values*. University of Chicago Press.

van der Veld, W. and W. E. Saris (2004). Separation of error, method effects, instability, and attitude strength. In W. E. Saris and P. M. Sniderman (Eds.), *Studies in Public Opinion: Attitudes, Nonattitudes, Measurement Error, and Change*. Princeton University Press.

Verba, S. (1995). The citizen as respondent: Sample surveys and american democracy presidential address. *The American Political Science Review 90*, 1–7.

Weßels, B. (2007). Political representation and democracy. In R. J. Dalton and H.-D. Klingemann (Eds.), *Oxford Handbook of Political Behavior*, Chapter 45, pp. 833–849. Oxford University Press.

Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University Press.

Zaller, J. and S. Feldman (1992). A simple theory of the survey response: Answering questions versus revealing preferences. *American Journal of Political Science 36*, 579–616.