Choosing an Identity

A General Model of Preference and Belief Formation

Sun-Ki Chai

Ann Arbor
For my family
## Contents

Preface vi

1. The Success and Failure of Rational Choice 1
   - The Assumptions of the Rational Choice Approach 5
   - The Strengths of Conventional Rational Choice 8
   - The Weaknesses of Conventional Rational Choice 12
   - The Chapter Structure of This Book 19

   - Structural Assumptions and Models 25
   - Assumptions about Decision Making 31
   - Preference and Belief Assumptions 59
   - Conclusion 77

3. A General Model of Preference and Belief Formation 81
   - An Identity Coherence Model of Preference and Belief Formation 82
   - Assumptions of the Model 83
   - Basic Implications of the Model 97
   - Preference and Belief Change 100
   - Rewards and Preference and Belief Change 105
   - Implications for Dynamic Choice 111
   - Implications for Collective Choice 116
   - Plan for the Following Chapters 124
   - Proofs of Theorems 124

4. Ideology Formation and Policy Choice in Ex-Colonies 129
   - Theories of Policy Formation 130
   - The Conventional Model of the State 133
   - Variations on the Conventional Model of the State 138
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.</td>
<td>The Origins of Ethnic Identity and Collective Action</td>
<td>174</td>
</tr>
<tr>
<td></td>
<td>Rationalist Theories of Ethnic Collective Action</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>A Coherence-Rational Choice Theory of Ethnic Group Formation</td>
<td>190</td>
</tr>
<tr>
<td></td>
<td>Case Studies</td>
<td>203</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>211</td>
</tr>
<tr>
<td>6.</td>
<td>Structural Change, Cultural Change, and Civic Violence</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>Theories of Tradition and Modernity</td>
<td>215</td>
</tr>
<tr>
<td></td>
<td>An Coherence Rational Choice Theory of Structural Change, Altruism, and Collective Action</td>
<td>228</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>241</td>
</tr>
<tr>
<td>7.</td>
<td>Conclusion</td>
<td>242</td>
</tr>
<tr>
<td></td>
<td>Modifying the Assumptions of the Model</td>
<td>242</td>
</tr>
<tr>
<td></td>
<td>Justifications for the Coherence Model</td>
<td>246</td>
</tr>
<tr>
<td></td>
<td>Summing Up</td>
<td>254</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>256</td>
</tr>
</tbody>
</table>
Preface

This book provides a general model of preference and belief formation, integrating it with a model of rationality to generate a unified model of preferences, beliefs, and actions. The basic concept behind the model is one that appears under a variety of guises, depending on the social science literature from which it is taken: regret, dissonance, and coherence to name but a few. I argue that there is an essential theoretical unity to these concepts and that, properly defined and constrained, they can form the basis of a general positive model with implications not only for rational choice theories, but also for issues of personal identity and culture. I also argue that identity and culture are not antithetical to rationality but instead are essential to its having any meaning. The central premise of the model is that individuals act to optimize their preferences and beliefs within a set of phenomenological constraints, analogously to the way that they act to optimize actions within a perceived set of structural constraints. Indeed, optimization is seen to occur jointly across preferences, beliefs, and actions, as individuals seek to construct an optimal life plan that constitutes their identities.

By providing this model, I hope to address some the major problems that have plagued attempts to extend the boundaries of the rational choice approach: (1) the ability to make determinate predictions and (2) the ability to make the transition from micro to macrolevel explanation. The model is examined in an extended fashion through three empirical studies that address major unresolved issues in the comparative study of long-term development.

While writing this book, I have incurred debts to a wide range of teachers and colleagues. This book was originally my dissertation at the Stanford political science department, and I would like to thank the members of my dissertation committee: my adviser Robert Packenham, Gabriel A. Almond, and John Ferejohn. I would also like to thank the people from whom I received regular comments and advice during various stages of writing: David Abernethy, Michael Hechter, James G. March, and the late Aaron Wildavsky.

I would also like to thank those who provided helpful comments on drafts of chapters, including George A. Akerlof, Jonathan Bendor, Richard Brody, Dennis Chong, William Dixon, Geoffrey Garrett, Kurt Gautatz, David Grusky, Ted R. Gurr, Dwight Hahn, Satoshi Kanazawa, Sunhyuk Kim, Masaru Kohno, Stephen D. Krasner, David Laitin, Hye-ryeon Lee, Linda D. Molm, Susan Olzak, Kenneth Organski, Bertrand Roehner, Tibor Scitovsky, Paul M. Sniderman, the late Amos Tversky, and Barry R. Weingast. I apologize to anyone whose name I have left out inadvertently. I also would like to thank the members of the Pizza and Politics and the Brown Bag Dissertation Writers groups at Stanford for risking indigestion on several presentations of my main ideas. I would also like to thank the University of Arizona sociology department for being ecumenical enough to hire a graduate from another discipline. Thanks also
to research assistants Jason Miller and Kristi Clark for help with copyediting the document.

I would like to express appreciation to everyone at the University of Michigan Press, including the successive politics editors: Malcolm Litchfield, Charles T. Myers, Jeremy Shine as well as their staff, for their admirable patience, as well as to two anonymous reviewers and one anonymous member of the executive committee for their detailed comments. Along with the usual thanks, I would also like to offer some apologies. First of all, to the people at Michigan, who were promised a completed manuscript by August 1995. This clearly did not happen! Sorry for the long wait, and I hope that the resulting improvements in the text at least partly compensate for it.

I should also acknowledge my indebtedness to the text-processing language Plain TeX, which was used to typeset this document, and the programming languages Snobol4 and Icon, which were used to write scripts for converting the citations, generating the index, and dealing with other last-minute changes of mind on formatting.

Moreover, as an aside to those scholars whose theories are discussed and compared in this book: the nature of my topic has meant that I have had to skate over an extremely wide range of literature in a relatively short space. Hence, the dangers of superficiality or misrepresentation are omnipresent. Despite my best efforts, I am sure I have not always managed to avoid such dangers. I should also emphasize that each theory is evaluated in light of its applicability to the particular issue being discussed in the relevant chapter, not with regard to its overall usefulness, and I hope my comments will be taken in that light. Nonetheless, to those whose theories I have wronged in some fashion, I offer apologies in advance.

Finally, I would like to apologize to my wife, Hye-ryeon Lee, and my son, Alex, for seeming to walk around in a fog at times while I was writing the book.
Chapter 1
The Success and Failure of Rational Choice

The rational choice approach, despite widespread criticism, has reached a point of unrivaled prominence among general theoretical approaches for explaining human action. This prominence extends across the entire range of social sciences. In economics, rational choice remains unchallenged as the dominant, if not defining, theoretical paradigm, and is sometimes referred to simply as the “economic approach.” ¹ In political science, largely under auspices of the public choice school, the rational choice approach has grown to the point where it has more adherents than any other, and its threatened dominance has set off an intense debate that has polarized the discipline. ² In psychology, rational choice can claim close ties to a wide variety of theories located under the broad rubric of expectancy-value analysis. ³ Furthermore, the rapidly-growing subfield of decision theory, while often rejecting the economist’s optimizing version of rational choice, uses it as the standard against which to compare its own bounded rational choice models. ⁴ In sociology, rational choice has risen from obscurity to become a major theoretical approach in both Europe and the United States, benefiting from the strong support of some the most prominent names in the discipline. ⁵ While anthropology is less interested than other social science disciplines in grand theories, rational choice has nonetheless been at the center of perhaps the main ongoing theoretical debate in cultural and social anthropology, that over whether interpretation of unfamiliar cultural practices should always proceed on the assumption that participants in such practices are rational. ⁶ Furthermore, it could

¹ See, for instance, Becker 1976b; Monroe 1991; Radnitzky and Bernholz 1987; Radnitzky 1992.
² Numerous books have been written on the topic of rational choice theory and political science, many of which are cited later in this chapter. For a journalist’s overview of this intradisciplinary strife, see Cushman 1994.
³ Although the exact boundaries of expectancy-value analysis are open to debate, it is usually defined as encompassing theories in which action is based on the value placed on different outcomes and the perceived effect of each alternative in achieving these outcomes. These include achievement motivation theory, some versions of social learning theory, some versions of decision theory, the theory of reasoned action, field theory, and purposive behavior theory. Although most of these theories do not formally specify their models of action, the major survey of the expectancy-value approach explicitly equates expectancy-value with expected utility, a key assumption in conventional rational choice theories. See Feather 1982, intro., 1.
⁴ For surveys, see Payne, Bettman, and Johnson 1992 and Abelson and Levi 1985. More citations are included in the section on decision theory in chapter 2.
⁵ Coleman 1990, which is arguably the most-discussed work in theoretical sociology to appear in the past decade, is dedicated to rebuilding the discipline on a base of rational choice assumptions.
Choosing an Identity

plausibly be argued that the de facto actor assumptions of most midrange theories in anthropology are rational choice assumptions, even if theorists are hesitant to use this label. In philosophy, rational choice plays the central role in subfields that border most closely on the social sciences. Many, if not most, of the debates in the contemporary philosophy of mind have revolved around rationality and its somewhat milder sibling, intentionality. Likewise, the most prominent theories of recent years in moral and political philosophy have been based explicitly on rational choice assumptions. Finally, the bulk of theoretical work in postwar philosophy of language has been devoted to specifying principles that rational individuals would use in communicating with one another. This theoretical focus is shared by pragmatics, the subfield of linguistics most concerned with human action. Furthermore, philosophy of mind, philosophy of language, and pragmatics all share cultural and social anthropology’s concern with whether actions (communicative or otherwise) should be interpreted under the assumption that the relevant actors are rational.

7 A prominent survey of anthropological theory notes that, “At the moment, the dominant theory of motivation in practice anthropology is derived from interest theory. The model is that of an essentially individualist, and somewhat aggressive actor, self-interested, rational, pragmatic, and perhaps with a maximizing orientation as well” (Ortner 1984, 151).

8 This literature is huge, and no volume covers in detail the full range of issues relevant to these debates. However, the following two survey books are to major parts of it: Stein 1996 concentrating on the empirical question of whether humans are rational, intertwining it with the normative question of what it means to be rational, and Lyons 1995 discusses different opinions that philosophers have on the proper role of intentionality in the explanation of human thought and action. Intentionality is generally defined as the description of individual states in terms of beliefs, desires and intentions toward action, with a systematic linkage made between the former two and the latter. Hence in social scientist’s terms it could be defined as encompassing both optimizing and bounded rationality.

9 Without arousing much controversy, it could be said that the three most influential works in moral and political philosophy since the 1970s have been Rawls 1971; Nozick 1974; and Gauthier 1986.

10 A popular recent textbook in pragmatics argues that the underlying assumption of philosophy of language and of pragmatics is “a rational language user,” and that their avowed aim is “to construct a rational philosophy of language use” (Mey 1993, 67). Another popular textbook argues that the mainstream (Gricean) theoretical maxims in pragmatics “describe rational means for conducting co-operative exchanges. If this is so, we would expect them to govern aspects of non-linguistic behavior as well, and indeed they seem to do so.” Hence, these maxims derive from “general considerations of rationality applying to all kinds of co-operative exchanges (author’s italics)” (Levinson 1983, 103).

11 The position favoring rationalizing interpretations is referred to in philosophy as the “principle of charity,” originating in Quine’s ideas on interpreting speech and extended by Davidson to a general position on interpreting all actions. See Quine 1960, 57-61 and Davidson 1984, chaps. 9-13. The most influential work of recent years in pragmatics was co-written by a cultural anthropologist and a philosopher of language. See Sperber and Wilson [1986] 1995, which presents a theory that is not only rationalistic but in effect adapts to interpersonal communication the common knowledge assumptions found in conventional rational choice theories of
Given all this, it can be said without great controversy that no other theoretical approach in this century has ever enjoyed the same level of ubiquity throughout the social sciences as the rational choice approach enjoys today. Despite this ubiquity, the success of the approach has been very tenuous. Its advance has been accompanied by an intense debate over its relative merit. The approach has been subject to the usual criticism of blatant inaccuracy given by outsiders to any would-be theoretical hegemon and to a predictable fuzzying of its assumptions as it is adapted to a wider and wider range of empirical phenomena. More notably, however, some of the most virulent mere criticisms of the rational choice approach have come from within its own ranks.

Despite the absence of any clear-cut resolution to this debate, there appears to be a growing consensus about the strengths and weaknesses of the approach. To put it briefly, the main strengths of the approach are that its conventional assumptions about actors are parsimonious and applicable to a very broad range of environments, generating usually falsifiable and sometimes empirically confirmed hypotheses about action across these environments. This provides conventional rational choice with a combination of generality and predictive power not found in other approaches. However, the conventional assumptions of rational choice not only lack verisimilitude in many circumstances but also fail to accurately predict a wide range of human behaviors.

It has become clear to many social scientists, including those who work within the approach, that fundamental changes need to be made to the conventional rational choice model. The critics agree that a modified model should be sensitive to cultural diversity and change yet retain generality and allow for deductive hypothesis-building.

This book presents a new model of preference and belief formation, showing how it can form the basis for making such changes. The model draws upon theoretical propositions and empirical generalizations from a wide range of disciplines, particularly from psychological theories of motivation and cognition, sociological theories of value and identity formation, and philosophical theories of coherence and higher-order rationality, integrating them into a unified set of assumptions. It will then be combined with a pared-down rationality-based model of decision making to form a general model of action.

In the substantive chapters of the book, the model of action will be used to analyze a number of major questions in the analysis of comparative development. What accounts for variations in economic development policy choices among Third World leaders? What factors determine the boundaries of ethnic identity in developing countries, and how do these boundaries affect collective political action? What is the causal relationship between socioeconomic change, political culture and levels of political violence, and how does it operate at an individual level? For each question, the hypotheses generated by the model
Choosing an Identity

are compared to those generated by existing rational choice and non-rational-choice theories.

Although the model is not restricted to the analysis of comparative development, this is the one major area of social science where conventional rational choice theories have been seen as most deficient, hence it provides the best opportunity to demonstrate how the new approach is different. Furthermore, it is an area where non-rational-choice theories have often been criticized for lacking generality or a clear deductive basis, exactly what rational choice could potentially provide.

This book is not meant to do several things. It is not designed to demonstrate that there exists a single, best model of preference and belief formation in all important respects, even from a purely positivist criteria. Hence, the book is not meant, by various refinements, to establish the theoretical hegemony of a particular version of the rational choice approach or to demonstrate its all-purpose superiority to other approaches. Instead, it is meant to show that it is possible to devise general assumptions about preference and belief based upon a broad integration of existing ideas in the social sciences and that these assumptions can explain variations between actors and changes within actors over time and significantly expand the predictive power of theories based upon rational choice models. It is also meant to show that a richer set of psychological factors can be systematically introduced into rational choice analysis without robbing it of its generality and deductive rigor. Finally, it is meant to show that a diverse set of social phenomena involving identity, ideology, and culture can be analyzed productively in a reasonably accurate manner using a parsimonious set of assumptions about thought and action.

A few notes about terminology: Throughout this book, I use the term actors to refer in a general fashion to the units endowed with agency by various versions of the rational choice model. Despite oft-professed methodological individualism, rational choice models have been applied to a wide range of aggregate units, including firms, interest groups, political parties, and states. Environment refers broadly to all relevant variables affecting outcomes, excluding the attitudinal characteristics of actors. Under this definition, an actor’s environment includes such things as his, her, or its own physical characteristics, location, and resource endowment. Attitudes are defined as the mental orientations that a particular actor holds toward other actors, the environment, and actions at a particular point in time. For rational choice models, the only attitudinal variables relevant to action are beliefs and preferences. Decision making refers to the mental process that determines action given a particular set of attitudes. Beliefs refer

---

\[I\] define attitudes here in the psychological sense of the term, that is, cognitions, motivations and evaluations. This is similar to what philosophers refer to as propositional attitudes, the most prominent of these being beliefs, desires, and intentions. The conventional rational choice model does to analyze evaluations/intentions separately from actions themselves, though this is not the case with some of the non-conventional models (e.g., time preferences) surveyed in chapter 2 or my own model which is presented in chapter 3.
to empirical cognitions, not to normative judgments. Preferences refer to the relative desirability of the outcomes generated by actions, not that of the actions themselves, unless an action is “expressive”, that is, valued intrinsically and hence treated as its own outcome. Preference and/or belief formation refers to a mental process that determines the values of these variables given the past and present experiences of an actor.

The term model refers to a set of assumptions about an abstract entity or process, such as a model of decision making or a model of preference and belief formation. Model of action or actor model refers to a combined set of assumptions about both decision making and attitude formation. The term theory, on the other hand, refers to a complete set of assumptions that is used to generate predictions under a specified set of scope conditions. Hence, a rational choice theory incorporates some kind of rational choice actor model, but by necessity also incorporate a set of structural assumptions, i.e., assumptions about environments. Ad hoc assumptions refer to assumptions in theories that are not explicitly derived from any model.

The Assumptions of the Rational Choice Approach
The cornerstone of the rational choice approach is what I refer to as the rational optimization model. Although rational optimization has been defined and formalized in many ways, in practice, most definitions are consistent with the following assumptions.

(1) All actors hold a set of logically consistent beliefs about the outcomes that will result from their actions. These beliefs may be deterministic, assigning a single possible outcome to each action, or nondeterministic. Nondeterministic beliefs may specify the probabilities of different outcomes for each action or merely identify some range of possible outcomes for each action. The conditions corresponding to holding each of these three types of beliefs are usually referred to as certainty, risk, and uncertainty, respectively. Probabilistic beliefs are assumed not to violate the usual axioms of probability theory. Each outcome is entirely a consequence of chosen actions and the existing state of the environment, implying that nondeterministic beliefs arise from an actor’s lack of knowledge about the environment or the choices of other actors.

(2) Actors have preferences that place the relative desirability of all outcomes in an unambiguous rank order. In other words, the usual requirements of a mathematical strict order, completeness/

---

13 Although it is not too uncommon for theorists to refer to “preferences” over actions, this usage is confusing, since it fails to differentiate between choices and the basis for those choices.

14 The three standard works credited with providing the formal foundations for the rational optimization model are Ramsey 1931; Von Neumann and Morgenstern 1944; and Savage 1953, though Ramsey has relatively greater influence among philosophers, and Von Neumann and Morgenstern among economists.

15 This quality is referred to as coherence in De Finetti 1980, 62-63. It differs from the epistemological concept of coherence discussed in chapter 3.
Choosing an Identity

connectedness, asymmetry, and transitivity, are satisfied by the preferred-to relationship. Preferences are representable by a utility function that assigns a number to each possible outcome, where such a number represents the magnitude of the actor’s desire for the outcome.

(3) Actors will optimize across actions given preferences and beliefs. When they have deterministic beliefs about the consequences resulting from each action, they will choose the action believed to lead to the outcome with the highest utility. When they have probabilistic beliefs, they will choose actions that maximize expected utility, i.e., the sum of the utility of each possible outcome weighted by its perceived probability. When they have nondeterministic, nonprobabilistic beliefs, they will choose actions that exhibit dominance over all other actions, that are believed to provide at least as much utility as all other actions under all possible states of the environment, provided such actions exist.

By itself, the rational optimization model cannot be used to make predictions about actions. While the first two components of the assumption essentially require that preferences and beliefs be compatible with an optimization procedure, they do not specify their content, leaving this exogenous to the resulting model. While there is broad uniformity within the rational choice approach in assumptions about decision making, there is more diversity in assumptions about the types of preferences and beliefs actors will have. On the other hand, a very large percentage of rational choice theories incorporates certain common types of assumptions about preferences and beliefs.

In most rational choice theories, beliefs are based on what can be called the information assumption. It posits that beliefs derive solely from observations of the environment and propositions that can be

---

16 The requirement of asymmetry is sometimes replaced with equivalent requirements of irreflexivity and acyclicity. By implication, the usual requirements of a weak order, i.e., completeness/connectedness, reflexivity, asymmetry, and transitivity, are thereby met for the “at least as preferable” relationship.

17 Such cardinal utility functions are equal-interval measures of desirability, which are necessary if a weighted sum of the utility of multiple outcomes, such as expected utility, will itself have the required characteristics of a preference relationship. The main uses of non-cardinal or ordinal utility can be found in cases in which the outcomes from each possible action are certain and do not involve sequences of states, hence such aggregation or cumulation is unnecessary. Indifference curves, for instance, represent ordinal rather than cardinal utility.

18 The definition of dominance provided here is that for weak dominance, which adopted by most versions of the independence assumption, and is in turn included in most specifications of the rational optimization assumption. Strict dominance requires that an actor believe that an action provides more utility than all other actions in all possible states of the environment. Whether dominance is required to be strict in order to ensure the choice of a particular action is largely irrelevant for applications of rational choice assumptions to real world phenomena. Preferences, unlike choices, can plausibly be assumed to exhibit fairly fine discrimination (though not infinitely so) over differences in outcomes. This in turn reduces to virtually zero probability the states in which an actor will be completely indifferent between two outcomes.
inferred from such observations. Observability in turn is a function of an actor’s location, material resources, and other characteristics of the environment, rather than of his, her, or its own mental characteristics. Nonprobabilistic beliefs are updated in light of observations through logical inference. Where beliefs are probabilistic, as is typical in models of incomplete information, it is usually held that an objective probability distribution for the relevant variable is known to each actor and that beliefs are updated through Bayesian inference. This implies that any difference in beliefs between actors must stem from differences in past observations. Actors are usually assigned either deterministic or probabilistic beliefs about their environment, which allows them to always calculate an optimal action under conditions involving unilateral action. On the other hand, they are rarely assigned deterministic or probabilistic beliefs that directly specify the actions of other actors. This leaves dominance as the main criterion for calculating optimality under conditions of strategic interaction, where outcomes depend on the simultaneous or mutually unobservable actions of multiple actors. Given the wide range of conditions under which no dominant action exists, models of strategic interaction tend to focus on identifying jointly optimal equilibria, profiles of strategies over actions for some group such that each member’s strategy is optimal given the strategy of the other members.

The tendency for rational choice theorists to apply different standards for beliefs about an environment and other actors reflects the view that the latter ought to be consistent with actors applying the rational optimization assumption to one another. Indeed, a typical addition to the information assumption for models of strategic interaction is the assumption of common knowledge of rationality. This is typically taken to mean that actors know that all other actors are rational optimizers, that all other actors know this, that all other actors know that all other actors know this, and so forth. Common knowledge allows actors to form beliefs about each others’ actions through iterated dominance calculations, in which actors rule out all present and future actions of others that never provide maximum utility, regardless of the state of the environment. Another kind of common knowledge assumption is usually included in rational choice models focusing on action over time, though this is rarely mentioned explicitly. This assumption states that actors know that their own future actions will be based upon rational optimization, that their future selves will be aware of this for all selves after them, and so forth.

Assumptions about preferences, unlike those about beliefs, fall into separate broad categories that nonetheless share certain major

19 For an attempt to reconcile strategic interaction with probabilistic beliefs by actors directly pertaining to each other’s future actions, see Skyrms 1992 and 1998. In Skyrms’s model, these beliefs are shaped through a process of iterated deliberation between actors which is itself strategic.

20 For a discussion of common knowledge of rationality, see for instance the symposium in the Journal of Economic Perspectives 6, 4 (fall 1992).
characteristics. First, for each actor utility reflects the quantity of a unidimensional, self-regarding maximand. In other words, all actors will maximize their own endowment of a single good. This does not mean that actors will never care about other goods, but such concerns will depend on beliefs linking the other goods to an increase in the maximand and will not be reflected directly in an actor’s utility function. Only the one good represented in a utility function is valued intrinsically, for its own sake, while other goods may be valued only extrinsically or instrumentally, as a means to an end. Furthermore, the maximand is isomorphic among all actors. In other words, although the utility functions of different actors are in conflict with one another (each actor cares about his, her, or its endowment of the good, not anybody else’s), they generate identical preferences for actors vis-à-vis their own relationship to the good. Where the outcome of an action is seen to extend over more than one future period, utility is calculated as the geometrically discounted cumulation of the quantity in all future periods.

Each category of preference assumptions is defined by its particular actor and maximand. Each category of preference assumptions, combined with the rational optimization assumption and the information assumption, generates a fully fleshed model of action for its designated actor. Each model in turn belongs to a major grand theory tradition in the social sciences. These include theories of market competition, in which actors are firms and the maximand is profit; theories of electoral competition, in which actors are political parties and the maximand is votes; and neorealist theories of international relations, where actors are states and the maximand is power. Most importantly, there is a general category that cuts across various rational choice traditions in which actors are individuals and the maximand is wealth or material consumption. These include theories of consumer choice and voting, which complement those of market and electoral competition; economic theories of labor markets and intrafirm relations; and political theories of collective action.

21 The distinction between intrinsic and extrinsic motivation is well-established in social psychology. See, for instance, Kruglanski 1975. In sociology, extrinsic preferences are usually referred to as instrumental. Hechter discusses immanent and instrumental preferences (1993, 216), while Lindenberg makes the similar distinction between universal and instrumental preferences (1990, 741). Both seem to have been influenced by Weber’s distinction between wertrational and zweckrational, with the latter generally translated as “instrumental rationality.” Among philosophers, Baier distinguishes between ultimate and preliminary ends (1958, 266-76). A related distinction often made by philosophers is that between desire simpliciter and prima facie desire. The latter describes desires that are inherent to objects themselves, rather being linked to properties believed to be possessed by them (Petit 1991, 151-59). In economics, the characteristics approach to consumer demand theory has sought to link demand to preferences over the characteristics of the products within consumption bundles rather than over the products themselves. This approach is associated with Gorman and Lancaster. See for instance Gorman 1956; 1980 and Lancaster 1966; 1971.
The Strengths of Conventional Rational Choice

I refer to these categories as forms of conventional rational choice. This may draw objections from those who claim that the rational choice approach is based solely on rational optimization and that no assumptions are made about preferences and beliefs. Nonetheless, whatever the virtues of this narrow view of rationality, assumptions about preferences and beliefs are required to make predictions about action. The alternative of inferring preferences and beliefs after the fact, from the actions being explained, makes the resulting theories tautological. Furthermore, the majority of rational choice theories that are applied to analyze specific empirical settings adopt some variety of the conventional assumptions described earlier.

One of the main advantages of conventional rational choice models is their generality. In other words, a single set of assumptions pertaining to each type of actor is compatible with any set of structural assumptions about the environmental setting in which the actor is present. Although this will certainly leave room for disagreement about the assumptions of specific theories, disagreement will be about the nature of the environment within the scope conditions set by the theory, not about the nature of the actors. This common deductive base provides a commensurability to rational choice theories, and sets them off from other theories that do not share their assumptions. By reducing the area for disagreement, this increases the amalgability of different theories into a single coherent body. Furthermore, it gives theories within the approach an a priori nature that protects them from accusations that they were constructed to “fit the facts” after outcomes were observed. Finally, it provides a base that can be used to generate midrange theories and hypotheses, even in analyzing settings where few systematic empirical data exist.

Another advantage of rational choice models is their parsimony. The information assumption, the common knowledge of rationality assumption, and the assumption of isomorphic and self-regarding utility function, when combined with the rational optimization model, allow rational choice theories to treat variations in choices among actors and by an actor over time as entirely a function of their structural position. To begin with, preferences and beliefs are seen as the only relevant attitudinal variables for determining action. Furthermore, rational choice models that leave preferences and beliefs exogenous are often referred to as “thin” models, see Elster 1983b, 1-2 and Ferejohn 1991, 282. Ferejohn uses the term thick rationality to refer to conventional rational choice, while Elster, in the same spirit, uses the term broad rationality. This common allusion arises from their basis in John Rawl’s earlier “thin theory of the good.” However, there is no terminological consensus on this usage. For instance, Michael Taylor uses the term thick to refer to egoistic rational actors, the difference with the other authors seeming to be that Taylor applies the terms to describe individuals, not theories (though he cites Elster in doing so). Moreover, the term thick is often associated in social science with Clifford Geertz’s concept of “thick description,” emphasizes local uniqueness rather than universality. See Rawls 1971, 396; Geertz 1974, chap. 1; Taylor 1989, 120; Taylor, 1988, intro., 66.

22 Rational choice models that leave preferences and beliefs exogenous are often referred to as “thin” models, see Elster 1983b, 1-2 and Ferejohn 1991, 282. Ferejohn uses the term thick rationality to refer to conventional rational choice, while Elster, in the same spirit, uses the term broad rationality. This common allusion arises from their basis in John Rawl’s earlier “thin theory of the good.” However, there is no terminological consensus on this usage. For instance, Michael Taylor uses the term thick to refer to egoistic rational actors, the difference with the other authors seeming to be that Taylor applies the terms to describe individuals, not theories (though he cites Elster in doing so). Moreover, the term thick is often associated in social science with Clifford Geertz’s concept of “thick description,” emphasizes local uniqueness rather than universality. See Rawls 1971, 396; Geertz 1974, chap. 1; Taylor 1989, 120; Taylor, 1988, intro., 66.

inference processes are uniformly based upon the axioms of formal logic and probability theory, which means that variations in beliefs about the environment are caused by variations in access to observable information. Utility functions are also uniform, which means that variations in preferences over outcomes are caused by variations in the effects of each outcome on an actor’s endowment of the maximand. Finally, actor beliefs about the preferences and beliefs of other actors are determined by their beliefs about each other actor’s structural position, since the inference processes, utility functions, and decision-making processes of all actors are taken to be uniform or isomorphic. Together, this implies that mental states need not be taken into consideration as independent variables because they merely intervene in a predictable manner between past and present states of environment and action.

The virtues of parsimony tend to be reinforced by generality. Generality ensures that different models of action need not be applied to different types of environments. This means that actor assumptions remain unified rather than splintering into a complicated mix of conditionals when theories based on the conventional rational choice model are combined. This in turn means that the parsimony of mid-range theories is better preserved when scope conditions are loosened and these theories are combined into a general covering law that can be used to predict actions in all environments.24

In addition to exhibiting generality and parsimony, conventional rational choice assumptions have displayed predictive advantages as well. The assumptions have been used to produce a wide variety of decisive theories, theories whose predictions regarding measurable real world phenomena rule out a much larger set of outcomes than what is already generally accepted to be impossible or unlikely.25 The narrowness of conventional rational choice actor assumptions ensures that the set of actions deemed possible for any actor is often fairly small given a reasonably complete set of structural assumptions. Of course, this implies that the decisiveness of rational choice theories depends on structural as well as actor assumptions. The nature of environments is the great unspecified area in the rational choice approach, and it is usually treated in an ad hoc manner in conventional rational choice theories. However, there are sufficient limits on the plausible ways in which social scientists can model environments to ensure that the resulting theories will have some real analytical bite. For instance, it is difficult to imagine a rational choice theory based upon

---

24 The notion of covering laws was first put forward in Hempel and Oppenheim 1948. The definitive discussion is in Hempel 1966.
25 Decisiveness can be seen as a hybrid of the concepts of falsifiability and determinacy. The trouble with these two concepts is that each is defined in either/or terms: falsifiability as the existence of even one conceivable set of empirical observations that could confirm a theory, determinacy as the existence of only one conceivable set that would not disconfirm it. Decisiveness would allow for a partial ordering, i.e., one theory could be seen as more decisive than another if the set of empirical observations it rules out is a proper superset of those ruled out by the other.
plausible structural assumptions in which vote-maximizing political parties choose positions that are opposed by virtually all voters. It is also difficult to imagine a model of collective political violence in which self-interested wealth-maximizing individuals will sacrifice their lives for a political cause.26

Another important factor in decisiveness is the extent to which the independent variables are amenable to reliable measurement. While the rational optimization model focuses on mental variables such as decision-making processes, preferences, and beliefs, conventional rational choice assumptions effectively do away with the need to measure these variables directly. Given the frequent difficulties associated with developing reliable indicators with reasonable face validity for mental variables, this view increases the ease with which conventional rational choice theories can be used to make decisive predictions vis-à-vis other theories that require measurement of mental variables. Furthermore, because of the compatibility of conventional rational choice actor assumptions with virtually all types of structural assumptions, conventional rational choice theories make use of similar structural variables to those found in structural theories, such as wealth distribution, technological development, military strength, and so on, and they can make use of the same indicators. At the same time, they explicitly and systematically specify a micro-level process through which a set of structural conditions leads to particular outcomes and not others.

In its combination of generality, parsimony, and decisiveness, conventional rational choice is clearly unmatched by any other set of widely used assumptions in social science. This would not be of much benefit if the predictions made by conventional rational choice theories had been universally disconfirmed by empirical evidence. While it is on this criterion of empirical accuracy that the rational choice approach has encountered its greatest criticism, it is also true that, on a wide range of issues, its theories have been accepted by a large segment of the social science community as being consistent with empirical evidence. Whether this acceptance is the product of rigorous testing against non-rational-choice theories is a hotly debated subject,27 but even a mixed record makes it difficult to dismiss the approach as inferior to others overall, given its strengths with respect to other criteria.

The perceived level of confirmation for conventional rational choice theories is, not surprisingly, the most widespread in economics, where the approach originated and remains dominant. Economic theories based upon conventional assumptions have long been used as the basis for both corporate economic forecasting, a use that presumably would not have survived if the predictions had been shown to differ greatly from actual measured outcomes. While such use in other disciplines

---

26 For a discussion of the role of plausibility in the adoption of structural assumptions and of the possibility for making such assumptions less ad hoc, see the beginning of chapter 2.

remains relatively rare, there has been for a long time an anticipation among proponents that this will occur.\textsuperscript{28}

Hence, the rational choice approach has been hailed by many as the prototype for a more rigorous, deductive and cumulative approach to political analysis. It has been called a “unified framework for understanding all human behavior,”\textsuperscript{29} the “most rigorous and the most general theory of social action that has been advanced in this century,”\textsuperscript{30} and the “universal grammar of social science.”\textsuperscript{31}

Proponents contrast its “hypo-deductive” nature with the mere “set of concepts and definitions” that comprise the functional and cultural approaches that have historically been its main methodological challengers.\textsuperscript{32} Others stress how rational choice can be applied to explain both micro (actor-level)\textsuperscript{33} and macro (environmental-level)\textsuperscript{34} phenomena, allowing a diverse set of insights to be inferred from a simple set of assumptions.\textsuperscript{35}

For these reasons, as well for others pertaining to the contemporary intellectual climate in the social sciences and in society as a whole,\textsuperscript{36} the rational choice approach has enjoyed a steady climb in prominence,\textsuperscript{37}

\textsuperscript{28} A pioneering early example of the use of rational choice assumptions for forecasting purposes in political science is Bueno de Mesquita, Newman, and Rabushka 1985.

\textsuperscript{29} Becker 1976a, 14. For a similar comment, see Lalman, Oppenheimer, and Swistak 1993, 98.

\textsuperscript{30} Rogowski 1997, 298, 304. See also Rogowski 1993, 443.

\textsuperscript{31} Hirshleifer 1985, 53.

\textsuperscript{32} Harsanyi 1969, 514. For similar views, see Mitchell 1969; Holt and Turner 1975; and Gray 1987.

\textsuperscript{33} Riker 1990, 174.

\textsuperscript{34} Coleman 1986, 360-63.

\textsuperscript{35} Friedman and Hechter 1990, 214-15; Harsanyi 1969, 515; Rogowski 1997, 300.

\textsuperscript{36} Among other things, the events and prevailing attitudes of the 1960s brought into disrepute any theory that seemed to imply that a basic normative consensus existed in any society or that differences in social outcomes are caused by cultural differences rather than differences in power. In addition, American social scientists were in large part imbued with a desire to emulate the hard sciences, with a corresponding proquantification ideology, and they could not object very strongly to methods adapted from the most mathematical of social sciences, economics. See Rogowski 1997, 304-6; Mansbridge 1990c, 9-11; and Alexander 1987, 114-18 for short discussions of such issues.

\textsuperscript{37} This is true of political science in particular. One indication is the number of accepted articles in the American Political Science Review categorized as “Formal Theory and Methodology,” which has increased by over 60 percent between 1985 and 1991. Since the rational choice approach dominates formal theory in political science, it is fair to interpret this as an indication of the increasing popularity of the approach. See Patterson, Bruce, and Crone 1991. In another measure, it was found to occupy about a quarter of all articles published in the American Political Science Review as well as the American Journal of Political Science. See Dow and Munger 1990, 607. For more on this increase, see Lahman, Oppenheimer, and Swistak 1993, 77.
a trend that has come be known as “economic imperialism.”

The Weaknesses of Conventional Rational Choice

This rise and the sometimes grandiose claims of the champions of rational choice have inevitably led many others to examine the approach’s weaknesses. Specifically, criticisms of conventional rational choice have centered both on the unrealistic nature of its assumptions and on its inability to predict a large number of important human phenomena.

Conventional rational choice assumes that utility functions are uni-dimensional (hence static), self-regarding, and isomorphic. Individuals in particular are assumed to maximize their own wealth. However, public opinion surveys, experiments, and ethnographies provide ample evidence suggesting individual preferences differ quite widely from this “ideal” of mono-maximization, incorporating such factors as altruism, expressive desires, and a sense of justice. Such studies moreover show a tendency for preferences to vary widely between individuals and to change significantly over time across populations. Although it is conceivable that this empirical evidence reflects an interaction between constant intrinsic utility and a changing environment, no one has been able to plausibly show how the types of preferences uncovered by these studies can be viewed as instrumental in maximizing the conventional utility functions.

Likewise, conventional rational choice assumes that beliefs arise purely from observable characteristics of the environment and propositions that can logically be deduced from them. However, this contradicts practically the entire empirical literature in social psychology, particularly that on decision making, attribution, motivation, and social cognition, much of which is devoted to showing that quite normal individuals are capable of holding beliefs that are not logically derived from and are even contrary to observable evidence. This empirical literature also shows how decision-making processes diverge widely from cool optimization, showing that a wide range of emotional factors come

38 See Swedberg 1990c, 13-33 and Swedberg 1990b; as well as Baron and Hannan 1995.
39 Broad general collections of such critiques include Hogarth and Reder 1986; Cook and Levi 1990; Nichols and Wright 1990; Monroe 1991; Coleman and Fararo 1992; Zey 1990; Ferber and Nelson 1993; Friedman 1996; a special issue of Political Psychology 16, 1 (March 1991); and a symposium in the Journal of Economic Literature 36, 1 (March 1998).
41 Among the most notable fairly recent works within a large literature are Almond and Verba 1980 and Inglehart 1990.
Choosing an Identity

into play and individuals will employ a wide range of heuristics, including those that exhibit a great deal of bias. Critics have also argued that both preferences and beliefs can be the results as well as the causes of actions, and that understanding the mutuality of this relationship is essential to understanding political, social and economic behavior. Emphasis on the endogenous nature of preference and belief change characterizes emerging approaches in a variety of subfields, including the “new institutionalist” approach to organizations, the “new social movements” approach to collective action, and the “reflectionist” approach to international relations. In response to these critics, defenders have turned to pragmatic arguments that conventional rational choice assumptions, regardless of their verisimilitude, can be justified as analytical tools because they are of great value in generating predictions about actions. According to such “as if” arguments, actors need not actually engage in conscious optimization, as long as the choices they make are similar to those they would make if they were doing so. As if arguments are bolstered by a priori justifications that point to the survival value of the designated maximands. The most popular takes the form of quasi-natural selection argument replicated within each of the rational choice grand theory traditions, which proceeds along the following lines: politicians need votes to remain politicians, firms need profits to remain firms, and sovereign states need military power to remain sovereign states. Those who fail to maximize the key unidimensional good will no longer exist, hence we can assume that all who remain maximize.

Such arguments are essentially functional, which is rather ironic given the polemics that are often launched by the rational choice camp

---

42 Because evidence for nonlogical or illogical beliefs and for nonoptimizing decision-making processes is found in practically every area of empirical social psychology, it is difficult to provide any definitive set of citations. Surveys of decision theory are a useful place to start, since they are the most likely to interpret their results in terms of their implications for the conventional rational choice model. See chapter 2. For a useful survey particularly aimed at economists, see Rabin 1998.


44 Rhoads 1990; Wilson 1980, editor’s conclusion.


47 For a review, see Keohane 1988. See also Lapid and Kratchowil 1995.

48 Originating in Friedman 1953.

49 For discussions of these imperatives in each category, see Mayhew 1974, chap. 1, esp. pp. 16-7 and Downs 1957, chap. 2, esp. pp. 30-31 for politicians and parties; Alchian 1950 for firms; and Waltz 1979, chap. 5, esp. pp. 88-93 for states. For a general defense of such justifications, see Ferejohn and Satz 1994. For an attack on their verisimilitude, see Green and Shapiro 1994, chap. 3.
against social science functionalism. There is no clear argument, other than from imitation, for how non-rational actors can adapt to pressures in the environment to behave in a rational fashion. Nor, for corporate actors based on aggregate units such as parties, firms and states, is there an explanation for how the interests of the varying individuals who comprise such units are reconciled in a way that allows the actor to behave in a maximizing fashion.

In each of these actor categories, moreover, the pressure of the environment will often be soft enough to leave a great deal of slack. Actors are seldom living on the edge of survival and need not maximize once they get beyond some reasonably achievable threshold. Hence plausible arguments can be made that political parties will pursue policy objectives as well as votes, firms will pursue market share as well as profits, and sovereign states will pursue wealth as well as military power. Natural selection arguments are even more difficult to make for the assumption that individuals maximize their wealth, since the survival threshold for wealth is so easily achieved in most contemporary societies as to make such an argument implausible. Moreover, there may be conditions under which acting to maximize personal wealth is self-defeating. If individuals can perceive the nature of each others’ utility function, it is quite possible that those individuals whose utility functions are known to include non-material goods can elicit better material treatment from others.

As-if arguments place prediction above explanation on the scientific set of priorities, and also view choices as the important dependent variables in action, rather than the processes by means of which actors arrive at such choices, priorities that are quite different from those explicitly held by ethnographers on the one hand and decision-theorists on the other. However, even if prediction of choice is given the top priority, critics point out that conventional rational choice theories are flawed not only because their assumptions are unrealistic but because they fail to predict a wide range of behavior. In fact, there is a wide range of analysis in which theories based upon conventional assumptions have been indecisive or disconfirmed by the evidence.

---

50 For comments on the links between evolutionary and functional arguments in social science, see Stinchcombe 1968, 85-99 and Thompson, Ellis, and Wildavsky 1990, 117-18, 201. For analysis of the use of functionalism in evolutionary biology, see Dupre 1987.

51 Hence the only prominent example that I know of where survival arguments have been plausibly made to support assumptions of individual economic welfare maximization (albeit extremely risk-averse) applies to subsistence agricultural societies. See Scott 1976, especially chapter 2.

52 See Frank 1988, chap. 5 and 6.

53 Moe 1979, for an effective rebuttal to Friedman that focuses in on crucial differences between the use of such assumptions in the natural sciences and in economics.

54 Simon 1979; 1986a.

55 The most systematic attack on the predictive power of conventional rational choice is Green and Shapiro 1994.
Even though decisiveness is seen as a relative strength of conventional rational choice, there remains a wide range of structural assumptions under which the model cannot provide decisive predictions. The most prominent example of such indecisiveness occurs in cases involving strategic interaction. Even if theorists seek only to identify an equilibrium profile of strategies over actions without predicting whether such equilibria will be reached, many types of structural assumptions will generate a large set of multiple equilibria or no equilibrium at all. This tendency is pointed out by the well-known “folk theorem” in noncooperative game theory, which shows that in most types of repeated interactions between actors in which the endpoint is nonexistent or impossible to estimate, just about any set of strategies over actions can be sustained as an equilibrium. Likewise, even interactions that take place only once will often generate multiple equilibria when actor payoffs are highly interdependent. In many real-world environments, it is difficult to assume that such a profile of strategies exists for individuals with conventional utility functions, given the extent to which the relative effect of different actions on every individual’s wealth is dependent on the actions of every other individual. Even if we disregard the strategies of other actors, other reasons for indecisiveness can be found in the immense complexities of the physical world and the difficulty of forming beliefs about causality from observable information and inference that are precise enough to isolate a single optimal action.

In real life, actors clearly make decisions based on beliefs about their environment that cannot be inferred solely from information, but draw instead from ideology and culture. Such beliefs are not antithetical to rationality. Indeed, in some cases they are essential to isolating one action from a myriad of possibilities and therefore necessary for rational decision making, something that has been noted by a wide range of thinkers from various social science disciplines.

While an increasing number of theories attempt to incorporate culture directly, rational choice theorists have more often added
ad hoc assumptions about beliefs and preferences on top of the conventional ones, without explaining their source. The addition of ad hoc assumptions, while permissible as a theory-construction device, removes much of the coherency and a priori nature from conventional rational choice assumptions about actors.

There is also a large and growing body of influential opinion that asserts that the conventional rational choice assumptions, particularly those about preferences, lead to inaccurate predictions about behavior. These criticisms extend across the entire range of conventional rational choice traditions and are backed up by a diverse collection of evidence about general patterns of behavior. More specific criticisms focus on paradoxes of both formal and informal participation in collective action despite the apparent lack of significant personal material incentives to do so. A huge body of work, much of it by rational choice theorists, has shown that it is necessary to modify conventional rational choice assumptions in order to explain why individuals vote in democracies and why they participate in social and political movements.

A prominent literature in political philosophy attacks the conventional assumptions for generating inadequate explanations for the functioning of democracies as well as for their effects on normative democratic theory. Similar criticisms have arisen in the field of international politics, where conventional assumptions about preferences and beliefs have been found to generate “empirical anomalies,” and in the study of domestic political competition, where theories assuming that parties seek solely stay in power and ignoring internalized ideology have been unable to explain a wide range of policy outcomes. Even among theories

---

59 Simon skewers a number of prominent rational choice theories, including much of voting theory, for just this reason. See Simon 1985 and Simon 1986b.


61 For general reviews of the evidence, see Etzioni 1988a, chap. 4; Lane 1991, chap. 17, 23; Mansbridge 1990c; Juster 1991, chap. 2-4; Schwartz 1993.


63 Green and Shapiro 1994, chap. 5; Chong 1991, chap. 1; Opp 1990, chap. 2; Muller and Opp 1986; Mason 1984; Buchanan 1979; Salter 1976, chap. 2; Silver 1974; Tullock 1971.

64 For two major works along these lines, see Mansbridge 1980 and Barber 1984. See also Brennan 1989; Petracca 1991 and Scalia 1991, as well as Barber 1993 and Mansbridge 1993.


66 See numerous cites in chapter 4.
Choosing an Identity

of the firm, where conventional rational choice assumptions originated, patterns of investment are often difficult to understand unless profit maximization is seen to take a back seat to other goals. The range of unexplainable behaviors is extremely large, and extends to every significant type of actor found in conventional rational choice theories and virtually every arena of action. In many environments, conventionally motivated actors would be predicted to act in a ludicrous manner, as Sen points out in this heartwarming scenario: “Where is the railway station?” he asks me. “There,” I say, pointing at the post office, “and would you please post this letter for me on the way?” “Yes,” he says, determined to open the envelope and check whether it contains something valuable.

Perhaps because of these explanatory weaknesses, theory building within conventional rational choice has gravitated toward increasingly complex and formal mathematical models of action in those realms, particularly formal economic and political institutions, where conventional assumptions are at least somewhat reasonable approximations of reality. By not accounting for the rich variety of human motivations and cognitions found in the world and focusing on predictable environments (particularly stable formal organizations), the conventional rational choice literature has directed research away from arenas where ideology, culture, identity, and related factors have a significant effect on actions.

This has had a particularly constricting effect on the study of comparative development. Although the conventional rational choice approach has been applied quite productively to Third World studies, this literature focuses primarily on bargaining and conflict among actors with prespecified interests and clear beliefs about institutional constraints rather than on the forces that create these interests and institutions. While the former are clearly important areas of analysis, neglect of other areas means ignoring of concerns that have traditionally been at the center of development theory, particularly comparative and long-term issues such as variations in the nature of social institutions across societies and the effect of economic development on political behavior.

---

68 Sen 1977, 332.
69 For a discussion of the effects of rational choice assumptions on narrowing the topics that are open for analysis, see Hirsch, Michaels, and Friedman 1987, 39-56.
70 Pye and Pye provide a disparaging, but not unrepresentative, evaluation of conventional rational choice formal theory by development specialists when they refer to it as “convoluted ways of elucidating the obvious by mathematical formulas.” See Pye and Pye 1985, 10.
71 By Ames, Bates, Geddes, Laitin, Levi, Popkin, Rogowski and Shirk, to name but a few in political science, as well as by numerous scholars in the field of development economics.
Such criticisms of conventional rational choice have built up to a grand chorus in recent years.\textsuperscript{74} and generated a growing literature that seeks to alter the conventional assumptions.\textsuperscript{75} However, attempts to reform conventional rational choice theory have encountered a number of obstacles. In particular, would-be reformers have faced the problem of changing assumptions to address empirical weaknesses without sacrificing the strengths of the conventional model. Although it has been relatively easy to find alternative sets of assumptions that can explain a particular set of phenomena better than conventional rational choice, it has been more difficult to generalize such assumptions to larger arenas without robbing the approach of its ability to make decisive predictions.\textsuperscript{76} Because of this, the use of unconventional assumptions has been characterized as adventures into “waters that are treacherous” and “very murky.”\textsuperscript{77}

\textsuperscript{74} There has been a cottage industry of books criticizing the usefulness of conventional rational choice assumptions even within its original domain of economics. These include direly entitled volumes such as Bell and Kristol \textsuperscript{1981}; Balough \textsuperscript{1982}; Wilber and Jameson \textsuperscript{1983}; Eichner \textsuperscript{1983}; Wiley and Routh \textsuperscript{1984}; Kuttner \textsuperscript{1984}; Artz \textsuperscript{1984}; Mirowski \textsuperscript{1989}; Rosenberg \textsuperscript{1992}; Brockway \textsuperscript{1993}; Ormerod \textsuperscript{1994}; Heilbroner and Milberg \textsuperscript{1996}. In addition, a number of recent presidents of the American Economic Association (Leontief, Gordon and Debreu, for instance) have taken to questioning the solidity of the methodological foundations of economic theory in their presidential addresses. A poll of economists taken several years ago showed that two-thirds felt that economics had lost its moorings. See Guzzardi \textsuperscript{1978}, cited in Maital \textsuperscript{1982}, 17. Although much work since then has been devoted to addressing shortcomings, it has not significantly altered the widespread perception both within and without the discipline that the approach as a whole is estranged from reality. For instance, see the fictional dialog between a management guru and an economist in The Economist \textsuperscript{325}, 7720 (December 21, 1991-January 3, 1992), 89-91.

\textsuperscript{75} An important manifestation of discontent with conventional rational choice is the effort by Etzioni and others to found a counter-hegemonic “socio-economic” approach. This has led to the creation of the Society for the Advancement of Socio-Economics, the Journal of Socio-Economics as well as a number of books, e.g., Etzioni \textsuperscript{1988a}; Etzioni and Lawrence \textsuperscript{1991} and Coughlin \textsuperscript{1991b}. Socio-Economics has since formed strong links with the existing approaches of behavioral economics and economic psychology, as manifested by the renaming the Journal of Behavioral Economics as the Journal of Socio-Economics in 1991 and the publication of articles by Etzioni on Socio-Economics in prominent edited collections on behavioral economics, economic psychology and in the Journal of Economic Psychology. The main difference between these labels is the emphasis that Etzioni puts on viewing socio-economics as a separate discipline from economics. See Etzioni \textsuperscript{1991}; \textsuperscript{1988b}; 1986.

\textsuperscript{76} For an argument that this has been a rather futile exercise, leading to theoretical “degeneration,” see Smelser \textsuperscript{1992b}.

\textsuperscript{77} See Coleman \textsuperscript{1990}, 292 and Hechter \textsuperscript{1994}, 320. Nonetheless, both Coleman and Hechter have been involved in pioneering efforts to develop new approaches to the issue of preferences. See Coleman \textsuperscript{1990}, chap. 13 and Hechter, Jasso, and Ranger-Moore \textsuperscript{1999}. To add to the turbulence of this analogy, the conventional rational choice assumptions of economics have themselves been characterized as “dangerous currents.” See Thurow \textsuperscript{1983}. 
Choosing an Identity

The Chapter Structure of This Book
This book attempts to venture into such waters, but it also tries to steer clear of some of the hazards involved. Chapter 2 surveys various alternatives to conventional rational choice. The first section of the chapter focuses on the effects of adding general structural assumptions to the conventional actor assumptions. The next few sections focus on ways to alter the conventional actor assumptions. Alternative decision making models include non-rational models based on norms; non-optimizing rational choice models of bounded rationality and satisficing, reinforcement, and non-expected utility, as well as normative philosophical models. Alternative general deductive models of preference include isomorphic utility functions, time preferences, consumption-based preference, and social exchange. I conclude that while each approach has certain advantages over conventional rational choice, none addresses the conventional approach’s main explanatory shortcomings while retaining its generality.

Chapter 3 presents a new model of preference and belief formation, using a formalized concept (“coherence”) that draws upon the psychological concept of dissonance and the economic concept of regret. The key assumption will be that individuals will, subject to certain simple constraints, repeatedly adjust their preferences and beliefs to minimize cumulative coherence, which is the subjective expected difference between the utility of the a posteriori best possible outcome and that of the outcome generated by actions chosen up to that point. This expected difference will be positive when the actor is not certain about future outcomes and prefers particular outcomes over others. I argue that this preference and belief model is consistent with the assumption of rational optimization, and combined with it significantly expands the explanatory range of the rational choice approach without sacrificing its strengths. The model is used to formally derive several theorems on how individual and group experiences generate preference and belief change over time, and these theorems are linked to a wide range of existing research in the social sciences. They cover social psychological phenomena such as bolstering, overjustification and reactance, as well such as wishful or unwishful thinking and “sour grapes” or “forbidden fruit.” Moreover, they go beyond individual processes to predict the conditions for the formation of altruism within groups and the acceptance of particular cultural norms.

Chapters 4, 5, and 6 contain substantive applications of the model. Each chapter examines a different major issue in comparative development: the origin of economic ideology among elites in developing countries and its effects on policies, the origins of ethnic boundaries and their effect on collective action, and finally the relationship between structural change, political culture, and political violence. In each chapter, I argue that a theory derived from the coherence rational choice model either helps shed new explanatory light on important empirical phenomena or provides greater deductive depth to existing explanations.

Chapter 4 attempts to account for the origins of economic ideology among the leaders of ex-colonial countries, as well as the way in which
this ideology has affected their economic development policy choices. I first examine existing positive theories of economic policy formation, focusing in particular on their explanations of variations in levels of state economic intervention among different countries. I argue that existing theories of policy formation either do not attempt to explain such variations, or do so in a way that ignores the key role of ideology in determining the extent to which states will intervene.

I then present a positive theory of ideology formation, using the process of coherence to construct a link between political conflict and the internalization of opposing ideologies by members of conflicting groups. This theory is used to explain the tendency among indigenous elites that engage in anticolonial activities to internalize ideologies that emphasize the economic benefits of policies diametrically opposed to those of the colonial power. It will be posited that such internalization will have a coherence-denhamming effect on elites who choose to fight for independence rather than collaborate, since it will increase the perceived benefits that will result if independence is achieved and hence reduce the risk that the decision was mistaken.

These internalized ideologies have a systematic effect on the rational policy choices made by such elites in their roles as postcolonial state leaders, including the extent to which they attempt to exert state control over the market. In particular it causes states in former Western colonies to intervene more heavily in their markets than those in non-colonies, even after other possible influences are taken into account. However, this effect is predicted to decline over time, as an increasingly large percentage of elites will be from a generation that has not participated in anticolonial struggles. A statistical test is conducted, regressing key explanatory variables isolated by this theory, along with alternative explanatory variables (e.g., GDP, population, land resources, and year-by-year changes in the global political environment), against four different measures of state economic intervention.

Chapter 5 focuses on the formation of ethnic identity and its relationship with ethnic collective action. I review existing general positive theories of ethnic collective action, including those theories that view such action as an opportunistic behavior on the part of self-interested individuals and those that attempt to incorporate ethnic identity directly into individual preferences and beliefs. I assert that former theories either cannot explain the existence of collective action based on ethnic criteria or attribute such action to various ethnic “resources” that are not fully conceptualized or operationalized. The latter theories use assumptions of ethnic preferences and beliefs to explain ethnic collective action, but do not explain the origins of these preferences and beliefs nor variations that exist in their nature and strength. I argue that no existing positive theory attempts to predict the location of the boundaries within which ethnic collective action takes place, nor the timing of their formation.

I then present a theory of ethnic group formation, one that is based upon links between cooperative action and coherence-reducing
Choosing an Identity

altruism. The theory is then applied to explain the formation of large-scale ethnic groups in developing countries, as well as the location of their boundaries. I first focus attention on an individual's formative experiences within societies which are just beginning to experience economic commercialization and large-scale political consolidation, hypothesizing that coherence-reducing processes related to cooperative action within a relatively stable and self-sufficient rural community create a sense of altruism towards other members of those relatively small communities. Later interactions within an urban metropole, however, lead to a broadening of the boundaries of rational cooperative action and the formation of larger groups. The location of these new boundaries are determined by an interaction between primordial altruistic ties of individuals to their local communities and their circumstantially shared interests with others in the economic and political environment, as well as by criteria of optimal size. In the long run, barring major changes in structural conditions, altruism will be redirected towards these new groups and large-scale ethnic identities will be formed.

A number of mini-case studies are presented to illustrate the theory, focusing on five of the most prominent cases of ethnicity “creation” that have occurred in the past century. These include the pan-Malay and pan-Igbo identity, the Luba-Kasai identity in Zaire, the Muhajir identity in Pakistan and the “local identity” in Hawai’i. In each case, the theory is used to explain why large-scale ethnic groups arose at a particular time in history and why ethnic cleavages formed around particular boundaries rather than others. While these case studies do not constitute a systematic test of the theory, they show the fruitfulness of the theory in making relatively accurate predictions across a wide range of historical circumstances.

Chapter 6 focuses on the relationship between socioeconomic change, political culture and civil violence. It begins by reviewing 19th and early 20th century theories of modernization that posited a linear movement from particularistic, collectivist “traditional communities” to impersonal “modern societies,” as well as how they been elaborated on by contemporary modernization theorists. I discuss various methodological and empirical critiques that have been directed at these theories, particularly allegations that the theories are unilinear, teleological and ethnocentric. I argue that not all modernization theories can accurately be described in those terms, and that recent theories in particular have broadened the narrow assumptions of their predecessors. However, in the process of doing so, so they have lost the common deductive basis that held the older theories together. I assert that such a deductive basis can only be restored if shared micro assumptions are brought into modernization theory.

I then present a theory, based on coherence model, that focuses on how changes in patterns of cooperative and conflictual interaction lead to changes in patterns of altruism, and how these in turn affect political behavior. Two types of change in interaction patterns are examined: the first, broadening, relates to the number of individuals with whom interactions take place. The second, fragmentation, relates
to the density of interactions with different groups of individuals. It is argued that broadening tends to lead to an increase in the scope of altruism, while fragmentation tends to lead to a decrease in its intensity. The theory is then applied to analyze the effects of long-term structural change, particularly changes in patterns of interaction that are created by such forces as economic commercialization and the consolidation of large-scale political units. Different patterns of such change lead to differences in the rate of broadening and fragmentation, which in turn lead to differences in altruism patterns.

These changes in cultural patterns in turn are shown to have systematic effects on rational political behavior, in particular on levels and types of political violence in society and the potential for authoritarian government. Broadening tends to initially increase the prospects for large-scale political violence (the advent of “plural-ethnic” society), then eventually decreasing them (the advent of “civil” society). Fragmentation tends to decrease the prospects of violence by reducing the amount of civil authority that is required to maintain order, but it also increases the prospects of successful authoritarianism. Unlike the other chapters, this chapter does not attempt to apply the theory to quantitative data or to specific cases. However, attempts are made to illustrate the processes with examples and to relate them with hypotheses found in the literature on political violence and the conditions for democracy.

Chapter 7, the conclusion of the book, discusses possible extensions and modifications of the coherence rational choice model. It also briefly discusses more general issues, including the integration of psychological and cultural explanations with economic explanations and the role of general actor models in social science theory.
Chapter 2
Alternatives to Conventional Rational Choice: A Survey

This chapter examines a wide range of existing alternatives to the conventional rational choice model. Because of the large number of approaches covered, the reviews are too brief to provide an overall analysis or critique of each. Instead, I attempt to evaluate each approach’s potential for remedying the major predictive shortcomings of conventional rational choice while retaining its strengths as a deductive, general, and predictive model. This involves analyzing the models found in each approach in terms of the main criteria described in the previous chapter: 
generality, the range of environments within which the model’s assumptions are applicable; 
parsimony, the overall simplicity of its assumptions; 
assertiveness, the extent to which predictions generated by the approach’s assumptions reduce the space of possible actions and outcomes in the environments that they apply to, and accuracy, the extent to which actions and outcomes that fall into the set deemed possible by the assumptions are confirmed by available empirical data. Each alternative approach will be judged by the extent to which it retains the conventional rational choice model’s generality, parsimony, and assertiveness, while improving overall accuracy and perhaps assertiveness as well.

While such criteria obviously imply certain value judgments about the purpose of social science theories, they should be seen as identifying a desirable constellation of characteristics, not as exhausting all possibilities. No attempt is made to address methodological debates over deduction versus induction, general versus mid-range theories or prediction versus explanation, other than to make the rather obvious point that different types of methodologies can serve complementary purposes. Indeed, although the rational choice approach is unambiguously deductive, general, and predictive, all versions of the rational choice model involve the explicit or implicit incorporation of insights from empirical findings as well as midrange and/or explanatory theories.

Besides the criteria mentioned, attention is also paid to an approach’s ability to explain the kinds of empirical indeterminacies and anomalies of social, political, and economic behavior discussed in chapter 1, without introducing new ones of its own. Indeterminacies often occur when consequences are long term and involve complex interactions between actors, with government economic planning and the formation of new large-scale institutions often cited as examples. Anomalies often occur in cases in which actors have a choice between action that is altruistic, discriminatory, or expressive and action that is self-interested. Voter turnout, participation in costly collective action, and ethnic group favoritism are some of the typical examples.
All social science theories contain two main components: environments (structure) and actors. Actors can be decomposed into a decision-making component and a preference and belief component (or, to use psychological terminology, a motivational and cognitive component). Hence, there are three main ways to reform the conventional rational choice approach through an alternative deductive model. The first is to build structural assumptions directly into the conventional rational choice model. The second to alter its assumptions about decision making, and the third is to alter conventional assumptions about preferences and/or beliefs. Each will be discussed in turn.

**Structural Assumptions and Models**

Despite its general assumptions about how actors make choices, conventional rational choice lacks general assumptions about the environments surrounding actors. Structural assumptions, however, are a necessary component of virtually any theory that seeks to predict action, since they specify the circumstances under which action takes place, including the position and resources of the various actors and the consequences attached to each actor’s choice. This relevance is particularly strong for rational choice theories. The rational optimization model implies that actors are constantly calculating the costs and benefits of their actions, and it is necessary to provide a specification of where these costs and benefits originate.

Although it is true that rational choice models can explain environmental conditions as an outcome of past actions, such recursive analysis cannot serve as a substitute for structural assumptions. Attempting to specify the state of the environment by analyzing past actions forces one to specify the environment of the past period, the period before that, and so on, eventually tracing things all the way back to the state of nature or primordial ooze. Even if we accept the dubious notion that this is practical, it would still require making general assumptions about the original state.

Because of the lack of general assumptions, structural assumptions are usually introduced into conventional rational choice theories in an ad hoc manner. Such assumptions typically specify the information that is posited to be available to actors, as well as the range of choices posited to be available to them. Without limits on the kinds of structural assumptions that are permitted, the environment is essentially a residual variable, one that can be manipulated to account for any behavior that might otherwise seem irrational.

In practice, the only limit on permissible structural assumptions has been the superficial plausibility of these assumptions. Proponents of pragmatic “as if” criteria for judging assumptions might argue that even plausibility is too strong of an a priori criterion, and can be ignored if the resulting theories have predictive value. However, in the absence of limitations on ad hoc structural assumptions, even conventional

---

1 For a similar argument, see Bicchieri 1993, 7-8.
2 For discussion along these lines, see Heap 1989, chapters 4 and 5.
rational choice theories would become unfalsifiable, since it would be extremely simple to develop implausible structural assumptions that would make any pattern of behavior optimal.\footnote{When explanations using conventional assumptions are difficult to generate, it is true that greater leeway is given in practice to implausible structural assumptions. Nonetheless, we still do not see a theory of voter turnout based on the assumption that there are leprechauns in the voting booths handing out money (more to educated and older people, less in primaries, etc), despite the fact that it arguably would require less stretching of the conventional actor assumptions than any rational choice theory of turnout that is currently available.}

If conventional rational choice is to be judged as a truly general model, it should be said to rule out a particular action only if this action is ruled out by by every permissible set of assumptions about the relevant environment. Equivalently, the set of outcomes deemed possible by the conventional rational choice model should be seen as the union of the sets of outcomes that are generated by all permissible sets of structural assumptions. One cannot claim that conventional rational choice assumptions have “predicted” a particular observed outcome when one particular set of structural assumptions plus the conventional rational actor model generate an assertive and accurate prediction, as long as substitution with another set of permissible structural assumptions does not lead to a similar prediction. In such a case, prediction is being generated in an ad hoc manner, and as much by the selected structural assumptions as by the conventional rational choice actor assumptions.

Hence, it is natural to investigate whether the assertiveness of conventional rational choice could be improved by replacing its ad hoc approach to structure with a more general set of structural assumptions. Possible candidates for such a set of structural assumptions are implicitly or explicitly associated with several of the prominent structural models in social science. For instance, neoclassical economics models assume that the only possible conditions in a free market are those in which the “invisible hand” will optimize consumer utility.\footnote{Sometimes a distinction is made between “formal”economics, which refers to the actor-oriented assumptions of rational choice, and “substantive”economics, which refers to structural assumptions of economics. This usage follows from Polanyi 1957, and was the basis for a major debate (primarily in economic anthropology) concerning which notions of economics were appropriate to apply to non-Western societies.} Substantive Marxist models assume that societies are divided between classes with great access to coercive economic power and classes with little such access, and hence the inevitable relationship is one of exploitation. Dependency theory models broaden these assumptions to an international context, with actors comprising a heterogeneous collection of classes, multinational corporations and states.\footnote{I am talking about those dependency theories that Packenham refers to as “holistic dependency,” where exploitation is seen as an unavoidable fact rather than a problem that can be addressed by action. See Packenham 1984.} Conflict theory models in sociology assume that the only possible arrangement of classes in society is two-level, with one class dominating access to coercive power in a variety of spheres (not only
Alternatives to Conventional Rational Choice: A Survey

Alternatives to Conventional Rational Choice: A Survey

economic), though change in access can occur. World systems theory models assume that the only possible arrangement of states in the world is into core, semiperiphery and periphery, with the core dominating access to production technology and capital, and where trade between core and periphery generally increases such gaps. Realist models in international politics assume that the state system is an anarchy in which survival depends on the principle of self-help.

Each of these approaches, of course, seeks to go beyond merely describing conditions in the environment to which it refers and attempts to predicting how these conditions shape the decisions of actors. Hence, each must have a set of actor-oriented (microlevel) assumptions attached to its structural (macrolevel) assumptions. And for each approach, the de facto assumptions adopted are compatible with those of conventional rational choice. Actors, whether they be states, social classes or individuals, are seen in these approaches to maximize their wealth in a calculating manner. Assumptions about beliefs are less standardized, but, except for occasional ad hoc propositions about false consciousness, the general implication seems to be that actors have complete relevant information about their environments (though the more pessimistic approaches imply that this information is of little use to them).

This may seem to make strange methodological bedfellows of structural Marxist, conflict, and dependency theories on the one hand and rational choice theories on the other. However, much of this seeming incompatibility stems from rational choice’s genealogical roots in neoclassical economics, where “invisible hand” structural assumptions have long dominated, and there is no logical reason why rational choice actor assumptions need be connected to neoclassical structural assumptions. Marxist materialism is quite compatible with conventional rational choice assumptions, and there is a growing literature on “rational choice Marxism.” Likewise, conflict theories, like rational choice theories and their close relatives in sociology, exchange theories, arose in large part to opposition to the normative emphasis of Parsonian structural-functionalism. Finally, the rise of both the dependency and rational choice approaches in comparative development were inspired by the belief that the modernization

---

6 For a discussion of this, see Munch and Smelser 1987, 380-85 and Boudon 1986, chap. 2. For criticisms of explanations that are advertised as “purely” structural, see Hechter 1992 and Taylor 1988, intro.

7 See related discussion in Marini 1992.

8 As Hechter notes, predominantly structural theories tend to predict that actors in similar circumstances will act similarly, hence they have a difficult time accommodating variations in preferences and belief. See Hechter 1992, 370.

9 Fukuyama 1989, 6 and Eisenstadt 1990, 249.

10 Prominent books include Roemer 1982; Elster 1985a and Przeworski 1985a. For overviews, see Roemer 1986 and Przeworski 1985b.

11 Alexander 1987, 163.
Choosing an Identity

approach entrenched the role of cultural norms and papered over the bedrock of “real” economic interests.\textsuperscript{12}

Whether neoclassical, Marxist or some other flavor, each of these sets of structural assumptions clearly restricts the possible ways the environments they describe can be modeled, hence increasing the coherency of theory-building within their associated approaches. However, each is quite a bit narrower, compared to the conventional rational choice actor assumptions, in the range of environments to which it is relevant. While rational choice actor assumptions can be applied to any environment in which humans, firms, political parties, or states are present, none of the above sets of structural assumptions can possibly be applied to anything but a small portion of these environments.

More general structural assumptions than the ones just mentioned exist within the functionalist tradition in sociology and anthropology, where structural variables are posited to be arranged in patterns that fulfill homeostatic, self-sustaining functions. However, because these assumptions are set at a very abstract level, they are generally not assertive and better suited for use as the basis for conceptual frameworks than as generators of falsifiable hypotheses.\textsuperscript{13} Rational choice theorists also sometimes implicitly adopt general functional structural assumptions when they argue that certain institutions exist because they generate efficient outcomes, without showing how rational action between actors can account for the formation of these institutions.\textsuperscript{14}

Although functional reasoning may be a starting point for an investigation of how certain structures come into being, the general assumption that structures fulfill certain abstract social functions (such as stability or efficiency) is rarely assertive because of the large variety of possible institutional arrangements that may serve such functions. Likewise, functional assumptions about structure are subject to the criticism that they fail to account for the empirical frequency of apparently dysfunctional collective outcomes, analogous to the criticism that optimizing arguments about action fail to account for the frequency of suboptimal choices.

In some cases, assumptions about the functionality of particular structural arrangements may be supported by quasi-evolutionary arguments linked to survival of the structures themselves,\textsuperscript{15} presumably within some even higher level metaenvironment, though this in effect turns structures into quasi-actors seeking to maximize their chances of survival. It is very difficult to use evolutionary arguments to support general structural assumptions that apply to entire societies or the entire world, since it is unclear exactly by what environmental factors such “actors” are being constrained. Without a clear specification of constraints, it is easy to come up with reasons why any arbitrary

\textsuperscript{12} See Bates 1990, 31-35 for rational choice and Smith 1985, 552 for dependency.

\textsuperscript{13} Homans 1964; Harsanyi 1969; Barry 1970, 166; Alexander 1987, 92.

\textsuperscript{14} See discussion in Granovetter 1985; Calvert 1996.

\textsuperscript{15} Stinchcombe 1990. Also see discussion in previous chapter.
structure might in fact be well-adapted to its environment, hence the frequent criticism that functional assumptions at such high levels often appear tautological when applied to real world structures.\footnote{See Alexander 1987, 107-10, Barry 1970, 169.}

One way to reduce this tendency toward tautology is by limiting the relevant environmental constraints to a single, readily measurable maximand. Then any structure that fails to achieve an optimal amount of the maximand for the individuals within it can be seen as to some extent maladaptive and hence, over the long term, anomalous. One approach that seeks to explain social structures in such a fashion is the cultural materialist approach in cultural anthropology. Here social structures are assumed to be selected for on the basis of their ability to facilitate the attainment of basic material needs by a society. Production and reproduction (procreation) are hence seen as the “infrastructure” upon which all other aspects of structure rest. The infrastructure is selected for its ability to provide for material needs. It in turn determines a corresponding “structure” of domestic and political economy, which in turn determines a “superstructure” of art, science, and recreation. These components of structure also determine a corresponding set of cultural preferences and beliefs.\footnote{See Harris 1979. Harris actually uses the anthropological terms etic and emic to refer to what I call structure and culture. The term “culture” is used more broadly in cultural anthropology to encompass both etic and emic.}

While cultural materialism manages to walk a tightrope in incorporating some of the assertive specificity of substantive structural approaches with the generality of structural functionalism, it still does not provide the level of assertiveness that would be required for its assumptions to have a great impact on the assertiveness of the rational choice model in predicting action.

Because environments consist of the analytical residual that remains after actors are taken out, structural models must deal with even more inherent heterogeneity than actor models, and it is difficult to think of structural assumptions that have non-trivial and falsifiable implications across more than a fraction of all possible environments. This implies that structural assumptions must likely remain relatively narrow in their scope in order to maintain their ability to generate assertive predictions about outcomes in the environments where they apply.

Without a general and assertive model of structure, the best alternative seems to be to formally classify different types of structural assumptions. Separating conventional rational choice into explicitly demarcated and clearly specified categories of dependency rational choice, Marxist rational choice and neoclassical economics rational
30 Choosing an Identity

choice would make it more difficult to manipulate structural assumptions to fit the phenomenon being studied.\textsuperscript{18} Hence, it would increase the assertiveness and consistency of theory-building within each category. This implies, in contrast to with the usual way of looking at things, that structural and actor-oriented assumptions are complementary, not competitive.\textsuperscript{19} The simultaneous existence of multiple categories of structural assumptions would not in itself prevent theoretical coherence, as long as each theorist does not mix and match from different categories of assumptions to suit the historical or geographical situation they are currently analyzing, but rather attempts to adopt a consistent set of assumptions across different situations within which the assumptions apply.

However, even within these categories of structural assumptions, a great deal about the environments described will remain unspecified; hence, none eliminates the need to add auxiliary ad hoc structural assumptions in order for the resulting mid-range theories to yield assertive predictions. Because of these methodological constraints, the informal criterion of plausibility applied to these ad hoc assumptions will remain an important ensurer of assertiveness within each category of structural assumptions. Although imprecise, plausibility (to the typical well-informed observer of the real world environment being modeled) must be maintained as a criterion for accepting or rejecting structural assumptions as long as general assumptions are not available or are incomplete.

The categorization of structural assumptions and the willingness of theorists to explicitly state the sources of the structural assumptions they adopt will undoubtedly be useful in promoting overall assertiveness. While the level of generality and assertiveness that can be attained will be limited, it will certainly be an improvement over the anarchy that currently reigns among rational choice views of environments. However, more coherent structural assumptions alone cannot solve the primary predictive shortcomings, discussed in chapter 1, for which conventional rational choice has been criticized. They do not provide assertive predictions when actors cannot plausibly be shown to have enough information to make optimizing choices. Nor do they explain numerous anomalies of action that cannot plausibly be shown to optimize an actor's wealth or corresponding self-regarding maximand.

\textsuperscript{18} For any such structural assumptions to be stable, however, it should not be rational for actors within the environments they describe to engage in actions that will transform the environment to make the assumptions false. This issue relates to the question of feedback loops and equilibrating mechanisms in functional theories, crucial for avoiding the much-criticized panglossianism that can otherwise plague functionalist theories. Merton 1948, chap. 1; Stinchcombe 1968, chap. 3, part II; Elster 1979, chap. 1; Thompson, Ellis, and Wildavsky 1990, chap. 11; and Eisenstadt 1990.

\textsuperscript{19} For a discussions on this point, see Kitschelt 1992; Taylor 1989. In practice, the reason why “structure-based” and “actor-based” theorists seem destined (structurally?) to be in competition has more to do with relative pride of place rather than theoretical incompatibility. It is the tendency of the former to imply that action is a fairly trivial matter once structure is determined, while the latter, as we have mentioned, tend to assign a residual role to structure.
Each of these predictive problems occurs because no plausible set of structural assumptions allows the conventional model to account for the outcomes being observed. Hence they cannot be solved by increasing general constraints on structural assumptions of rational choice theories, as desirable as such a reform might be nonetheless.

**Assumptions about Decision Making**

If adding general structural assumptions do not provide a complete solution to the main predictive shortcomings of the conventional rational choice model, then another possibility is to alter its assumptions about actors. This can be done either by providing alternative assumptions about decision making or by providing alternative assumptions about preferences and beliefs. Clearly, nearly all theoretical traditions in the social sciences have some relevance to the issues of decision making, preferences, and beliefs. However, I focus on models that have been specified in a way that is directly adaptable to the framework of the rational choice approach, either through some form of choice-theoretic specification or through direct incorporation into a nonconventional rational choice model of action.

For the most part, the nonconventional actor assumptions discussed in the rational choice literature focus on the individual as their unit of analysis. One reason for this is that it is very difficult to model divergences from rational optimization (i.e., bounded rationality or nonrationality) and changes in preferences and beliefs among higher-level actors without straining towards anthropomorphism. Theories that modify conventional assumptions tend to incorporate such factors as motivated biases, altruism, and addiction, characteristics that are more difficult to attribute to firms, parties, or states than to individuals. Another reason, however, is the tendency within the rational choice approach to give precedence to individual-level theories when they generate different predictions than higher-level theories do. While each category of conventional rational choice theory is based on a consistent set of actor assumptions, the union of these categories is not, and predictions generated by individual-level conventional rational choice theories may go beyond or even contradict predictions generated by conventional theories that take higher-level units as their actors. Individuals seeking to maximize their wealth will not necessarily on aggregate act to maximize profit, votes, power, or some other self-regarding good for the collective entity they belong to. Within the theory of the firm, this manifested most clearly in the debates about X-inefficiency, while the closest analogue to this in international relations are arguments over the importance of “bureaucratic politics” and of intrastate politics in general. In the

---

21 As made prominent in Allison 1971, chap. 5; and Halperin 1974.
22 For discussions on the various possible lower-unit level sources of a state’s preferences, see Nye 1988, 238-39, 248-49 and Jervis 1988, 324-29.
study of electoral competition, these debates revolve around the relative importance of intraparty politics.\textsuperscript{23}

Perhaps because of the general acceptance of methodological individualism among rational choice theorists, rational choice theorizing at higher levels of analysis is usually seen as a simplification that should give way to individual-level analysis if it displays major predictive shortcomings.\textsuperscript{24} Hence, the initial method for dealing with empirical indeterminacies and anomalies is usually to lower the level of analysis when this is possible. It is only once this is done and problems remain unresolved that attempts are made to alter actor assumptions. It is clear from earlier discussion, however, that important indeterminacies and anomalies remain in conventional rational choice even when it is taken down to the individual level.

The first way to alter the conventional rational choice actor model is to abandon the rational optimization model and replace it with an alternative set of assumptions about how preferences and beliefs determine actions for decision making. Alternative general assumptions about decision-making processes can be divided into two categories.

First, nonrationality assumptions abandon the idea that action is in any way goal oriented, as well as the concept of preferences. The most radical sets of such assumptions also abandon the idea that action is consciously directed, as well as the concept of beliefs. However, while there has been considerable discussion about the possibility of developing nonrational actor models by abandoning mental variables altogether or focusing on the primacy of norms, theorists have largely refrained from attempts to build general predictive models of this sort.

Second, nonoptimizing rationality assumptions preserve the goal-oriented nature of actors. They imply that actors seek to maximize utility in some fashion, but that they do not always choose optimal actions. A wide variety of decision-making models have been developed in this genre, primarily in psychology, but also including major bodies of work in philosophy and economics. Some of these models are explicitly defined as representing bounded rationality, or decision making that falls short of the ideal due to limitations in the mental capabilities of actors, but others are portrayed as representing a kind of decision making that is equally rational or more so than optimization.

Nonrationality Assumptions
The defining characteristic of any model of rationality, bounded or not, is that it assigns a significant causal role to both preferences and beliefs in determining action, either as independent or intervening variables. Therefore, the most radical alternative to rational choice assumptions is to drop from actor models all references to both preferences and beliefs, as well as all other mental variables.

\textsuperscript{23} See Aldrich 1983; Tsebelis 1990, chap. 5; Laver and Shepsle 1990; Gaines and Garrett 1993.

\textsuperscript{24} Indeed, the main criticism initially leveled by rational choice collective action theorists against earlier interest group theories was not that the latter failed to apply optimizing assumptions, but rather that they failed to apply them at an individual level.
Although the assumption of nonrationality clearly rules out decision making based upon the calculated pursuit of utility, it does not leave the theorist with any obvious alternative basis for an assertive model. In fact, it can create the initial impression that decisions are erratic or even random. After all, how can an individual make predictable decisions if these are not based on his or her own preferences and beliefs?

There are two major alternatives to the use of mental variables in the prediction of action. The first is to simply posit a direct causal link between environments and actions, bypassing the notion of agency altogether. The other is to fill the intervening role of mental variables with other variables linked to the internal states of actors, particularly physical variables. The first approach is associated primarily with behaviorist thought in psychology, while the second is associated primarily with connectionist thought in cognitive science. Arguably, these kinds of actor models should not be seen as modifying assumptions about decision making as much as dropping the whole distinction between attitude formation and decision making in favor of a more holistic views of action.

**Behaviorism**
The behaviorist approach to modeling action, which dominated much of American psychology until the 1950s\(^\text{25}\) and had a major influence on the fields of philosophy\(^\text{26}\) and sociology,\(^\text{27}\) and more indirectly on political

\(^{25}\) The manifesto of the behaviorist movement in psychology was Watson 1913. Comprehensive statements of methodological principles are presented in Watson 1919 and Watson 1930. The most prominent behaviorist of the latter half the century was undoubtably Skinner. Skinner’s model of action is described in Skinner 1953, chapters 17-8 and Skinner 1974, chapters 1-2. Skinner refers to his position as “radical behaviorism,” contrasting it to the “methodological behaviorism” emphasized by Watson and other early behaviorists, denying the reality of attitudinal states as well as the usefulness of using them as theoretical variables. However, his position is arguably (though somewhat ambiguously) less dogmatic than that of the early behaviorists, since he explicitly accepts the usefulness of making inferences about “private events,” even when they are not directly observable, though he rejects the idea that such events are attitudinal or can be known through introspection. See Skinner 1953, 281-82.

\(^{26}\) Behaviorism was for a long time considered one major approach to the philosophy of mind; the most influential work in this respect is Ryle 1949. It was also closely associated with the logical positivist approach in the philosophy of science, most explicitly in the work of Carnap and later Hempel. See Smith 1986b for a broad study of the subject.

\(^{27}\) The founders of the exchange theory approach in sociology were greatly influenced by the reinforcement-based ideas of behaviorism as well as by rational choice theories; see section on social exchange models of value later in this chapter.
science and economics has long maintained that mental variables such as preferences and beliefs should be left out of scientific theories. Instead, behaviorist actor models draw direct causal links between past and present environmental conditions, particularly contingencies of reinforcement and action. The conceptual distinction between actor and environment is thus abandoned, with all observable characteristics of individuals other than behavior itself being seen merely as part of the set of environmental conditions in which behavior takes place.

Aside from an a priori opposition to assigning causal force to non-physical and unobservable entities, two main pragmatic methodological justifications are forward for the behaviorist stance. The first is that mental variables, particularly attitudinal variables such as preferences and beliefs, are not subject to precise measurement and that including them in models will eliminate the ability to make assertive predictions. The issue of whether or not precise measurement of attitudinal variables is possible is of course a highly controversial and important one. However, it is largely irrelevant to the comparison between the behaviorist and conventional rational choice approaches. As was discussed in chapter 1, while the conventional rational choice model views preferences and beliefs as the direct causes of actions, they are rarely measured. Instead, (extrinsic) preferences and beliefs are seen as intervening between past and present structural influences and action. The conventional rational choice model reflects the causal relationship $\text{environment} \rightarrow \text{preferences/beliefs} \rightarrow \text{actions}$, where the first arrow represents a preference and belief formation process and the second a decision-making process, while a behavioral model reflects the causal relationship $\text{environment} \rightarrow \text{actions}$, where the arrow represents a reinforcement process. Hence, independent variables for the two types of model are the same and any conventional rational choice theory can clearly be converted to one that is effectively behaviorist, without changing the hypotheses it generates, by simply bypassing its description of intervening mental states.

Another possible justification for behaviorism is that the introduction of mental variables reduces the parsimony of actor models. This may seem to be the case, since inserting minds between environments and actions increases both the number of variables to consider and the number of links in the causal chain. However, such a judgment is premature because it does not consider the complexity of the specified relationships between variables. A rational choice model uses mental

---

28 Political science experienced a “behavioral revolution” in the 1960s, which aimed at redirecting study away from laws and formal institutions and towards patterns of political behavior, though this was influenced more by the psychological behaviorists’ notion of an empirically-based science of behavior rather than their injunction against the use of mental variables. The position of political behaviorists is summarized in two works by Eulau 1963 and Eulau 1969.

29 Behavioral economics, arising largely from the ideas of Simon, also adopts the psychological behaviorist’s emphasis on empirical verification, but explicitly incorporates mental variables into its models of boundedly rational decision making; see section on bounded rationality and satisficing later in this chapter.

states as a kind of causal bridging device, decomposing the overall process that determines action into separate processes of preference and belief formation and decision making. Such an indirect causal chain will simplify rather than complicate the formal representation of the actor model when preferences and beliefs are seen as being determined by a repeated process that takes place over multiple past periods.\footnote{Formally, take the behaviorist model $h: \text{environment} \rightarrow \text{actions}$, where $h$ is some mathematical function associated with a reinforcement process. Assume $h$ can be decomposed into functions $f$ and $g$ such that $h(e_0, e_1, \ldots, e_t) = g(f(e_0), f(e_1), \ldots, f(e_t))$, where each $e$ can be taken to represent a past environmental state. In such a case, specifying the function in two parts will avoid the need to write out the formula for $f$ separately for each argument $e_1, e_2, \ldots, e_t$. Suppose furthermore that the $f$’s correspond in some way to the concepts that we may refer to as changes in attitudes, or more specifically changes in preferences and/or beliefs. Then we can transform the behaviorist model to a more parsimonious rational choice model $f: \text{preferences/beliefs} \rightarrow \text{actions}$, where $f$ stands for the function associated with the preference and belief formation process and $g$ for the function associated with the decision-making process, and $h = f \cdot g$. In the conventional rational choice model, belief formation based upon information can be represented as a repeated process, either on observations or on time periods. The preference formation models discussed below do the same for utility.}

The behaviorist approach has not spawned many attempts at a general model of action, focusing more on incorporating laboratory findings into midrange theories about particular kinds of choice sets. The conjunction of such midrange theories would clearly make for a very complicated model, even if disagreements between them could be resolved. Moreover, it is notable that virtually every prominent behaviorist model that applies across multiple environments includes a variable that is functionally equivalent to utility, namely, the propensity of a specific reinforcer to elicit behavior. Indeed, the most prominent general model spawned by the behaviorist literature, Herrnstein’s matching law/melioration model, has over time rejected the behaviorist methodological stance by explicitly incorporating the concept of utility, which is why it is discussed later in this chapter in the section on bounded rationality. Furthermore, the behaviorism-influenced social exchange models of Homans and Emerson not only incorporate the concept of values, they attempt to account for the way in which values are formed, which accounts for their inclusion in the section on preference formation. Such prominent examples suggest that it is difficult to avoid modeling the internal states of actors, or close substitutes, when formulating general theories.

**Connectionism**

Connectionism refers to broadly to the movement in cognitive and computer science that seeks to develop computing systems based upon
the highly decentralized and parallel architecture of the human brain.\textsuperscript{32} The goals of connectionists are quite varied, and much work in this area is aimed less at modeling human thought or action than at generating practical analytical tools that can be applied to analyzing a wide variety of complex phenomena. However, it has also been associated with a number of prominent theorists, particularly among philosophers of mind, who have predicted and advocated the eventual replacement of mental variables in the analysis of human thought and action with physical variables, particularly those representing the characteristics of neurons and the synapses connecting them.\textsuperscript{33}

This “eliminativist” position\textsuperscript{34} has generated a major debate in cognitive science, since mental variables are central not only to rational choice theories in the social sciences, but also to traditional artificial intelligence in computer science, which bases its representation of knowledge on symbolic logic,\textsuperscript{35} and intentional analysis of action in philosophy, which focuses on the desires, beliefs, intentions, and other attitudes of individuals. Intentional analysis has in turn been associated by philosophers with the “folk psychology” that individuals use to make sense of one another in everyday life.\textsuperscript{36} This may seem to make rather strange bedfellows of formal rational choice models in social science and common sense folk psychology, but, similarly to the comparison with Marxist models, any seeming incompatibility has more to do with the genealogical roots of the former in neoclassical economics, and an inherited propensity for mathematical modeling, than with anything inherently mathematical about rational choice actor assumptions.

Eliminativist advocates of connectionism take the position that all action is completely determined by physical causes, and that decision making is therefore a neural process. Furthermore, they argue that there is no reason to believe that the mental concepts typically used in philosophical and social science actor models correspond to the

\textsuperscript{32} Although connectionist ideas have long been discussed in computer science, the work which triggered a wave of renewed interest is Rumelhart, McClelland, and PDP Research Group 1986. The terms neural networks and parallel distributed processing are often used interchangeably with that of connectionism, though there are some differences in connotation. In practice, any piece software based upon large numbers of relatively simple processes communicating with another without a supervising process on controlling them can be called connectionist, even when they are implemented on conventional non parallel hardware systems.

\textsuperscript{33} For a collection of articles on the subject, see Ramsey, Stich, and Rumelhart 1991. Some of the most influential of such arguments have been associated with the work of the Churchlands. See Churchland 1981; 1979; 1984; 1989 and Churchland 1986.

\textsuperscript{34} It is sometimes referred to more specifically as the eliminative materialist position to distinguish it from the earlier eliminative behaviorism discussed in the previous section.

\textsuperscript{35} A good overview of the respective positions of logicists and connectionists can be found in Nilsson 1991 and the response, Birnbaum 1991.

\textsuperscript{36} For a debate by major names in the field on eliminativism, see Greenwood 1991. For a connectionist broadside against folk psychology, see Stich 1983.
physical characteristics of neural states. Therefore, any theory based upon such concepts will inevitably be erroneous. Finally, they argue that recent advances in neuropsychology have opened the door to directly measuring these characteristics, and hence to their eventual incorporation into actor models at the expense of preferences, beliefs, and so forth.

Most of the philosophical debate over connectionist models deals with their ultimate status as a replacement for existing actor models, rather than with whether such a replacement is currently feasible. On the other hand, current feasibility is a central issue for social scientists. By this criterion, and despite its great advances, it is clear that neuropsychology has not reached the point where an actor model based primarily upon assumptions about neural processes can be used to generate assertive and accurate predictions about action across a broad range of plausible structural assumptions. The complexities of neural processes are simply too overwhelming at this point to draw clear links from environments to neural states and from neural states to actions in any general and assertive fashion.

It is true that connectionist computing systems, particularly those based upon learning algorithms such as backpropagation, have often been applied to predict actions or outcomes arising from collections of actions, most notably movements in financial markets. However, such systems cannot be said to constitute or form the basis for general and assertive actor models. Their main parameters, connection weights, are determined inductively by processing a large “training set” of data on past conditions and outcomes within a particular type of environment. Therefore, assertive predictions are limited to the environmental conditions described by the training set.

Furthermore, connectionist systems that are applied to real world social phenomena tend to be designed without any serious attempt to base the abstract nodes and connections of their algorithms upon physical characteristics of actual human brains other than their basic parallel structure. Hence they provide little evidence one way or the other on the superiority of basing social science theories on decision-making models with neurological verisimilitude. Instead, applied connectionist systems are often only incrementally different in the way they process inputs and generate outputs from traditional prediction, control, and pattern recognition systems based on inductive logical or statistical methods.

---

37 Skinner used similar reasons over four decades ago to reject the idea that neuropsychological variables had much of a role to play in predicting behavior. Arguably, the higher current level of knowledge about neural processes does not change the validity of this assessment. Skinner 1953, 27-29.


39 For instance comparative overviews, see Kulikowski and Weiss 1990; Schalkoff 1992, where “structural” refers to quasi-syntactic classification schemes; Michie, Spiegelhalter, and Taylor 1994; and Chekassky, Friedman, and Wechsler 1994.
that the most popular types of neural network algorithms are simply a form of statistical analysis.\textsuperscript{40}

Although it is more difficult to refute eliminativism independent of any time frame, one could plausibly argue that the inclusion of mental variables is necessary for building general, assertive, and accurate theories, or that it is necessary for expressing causal connections in a tractably parsimonious way.

The first position can be defended by taking the view that mental variables are not reducible to physical ones. This requires either adopting some sort of Cartesian mind-body dualism or differentiating between reducibility in specific cases and that which applies more generally. In philosophy of mind, the latter is identified with Davidson’s notion of “anomalous monism,” which accepts that the mental state of a particular individual at a particular point in time supervenes on, that is, completely determined by, some underlying neural state, but also posits that the characteristics of mental states cannot be systematically reduced to the characteristics of neural states.\textsuperscript{41} Given this, it is also necessary to show that the irreducible aspects of mental variables are essential rather than superfluous for analyzing action.\textsuperscript{42} Davidson argues that the characterization of mental states and actions is holistic, which in turn implies that actions are not identified independently of the mental states that are thought to cause them.\textsuperscript{43}

This is relevant for social science theories aimed at predicting action, since choice sets are nearly always defined using verbs that presuppose particular social objectives and knowledge, such as “buying” and “voting,” rather than in terms of the physical displacement of particular biological masses. Thus, choice sets have intentionality built into them. While this could be used as eliminationists as an argument for changing the way we define choice sets, we could not do so without fundamentally changing our linguistic conventions. For social scientists, this would also mean changing

\textsuperscript{40} See for instance Cheng and Titterington 1994 and White 1989.

\textsuperscript{41} Monism refers to the view that all reality is made up of a single substance, one that rejects a dualistic view of mind and body. Anti-dualism is more commonly referred to as materialism or physicalism, though the latter terms are more explicit about what the single substance is. Anomalous monism is usually contrasted with the “identity theory,” which posits that the mental can be reduced to the physical in all aspects, or is referred to as a theory of “token-token” rather than “type-type” identity. See Davidson 1980, chapters 11-13.

\textsuperscript{42} Indeed, it has often been argued that irreducibility is due to the fact that mental variables contain causally irrelevant information, which then implies that they have no role in scientific theories. Such variables are referred to as “wide” or “relational” because of their purported inclusion of information about the relationship of the individual to her environment. This argument is made most often in reference to the “twin earth” thought experiment describing a case in which individuals with identical neural configurations would nonetheless be said to have different mental states. The experiment was in turn used originally within the context of the externalism/internalism debate in the philosophy of language, over whether the meanings of words could be understood without knowledge of the relationship between the speaker and her environment. See Putnam 1975.

\textsuperscript{43} See Davidson 1980 and Davidson 1982.
our dependent variable rather than showing that mental variables are superfluous for what we are currently trying to predict.

The second position can also be linked to philosophical defenses of intentional analysis. For instance, Dennett defends intentional analysis on instrumental grounds, arguing that it provides the most effective way for humans to make predictions about action given the complexity of physical causation and their own bounded rationality.\footnote{These ideas were originally presented in Dennett 1971, chap. 1. They are developed in detail in Dennett 1989.} In a somewhat different vein, Gordon and Goldman argue that folk psychology is not a scientific theory but rather a kind of simulation in which individuals use their empathetic abilities to apply their own decision-making processes in predicting and making sense of each other’s actions.\footnote{Such positions are known as “simulation theories,” as opposed to “theory theories.” Whether simulation could itself be seen as theoretical would depend on a number of issues, including whether theories must be consciously held and be thought by their holders to represent actual processes. See Gordon 1986; 1992; Goldman 1989; 1992.} A modified version of such defenses could be used to suggest that, even if connectionist theories are potentially more general, assertive and accurate than those containing mental variables, it may be inherently more difficult to express such theories in a parsimonious fashion.

In order to use this to argue for retaining mental variables in social science theories, we need to show that relative tractability will not be altered by advances in scientific knowledge. The earlier discussion of behaviorism is thus illustrative. It is clear that no rational choice model can exceed the best behavioral model in generality, assertiveness and accuracy since any rational choice model can be transformed into a behaviorist one by merging the causal relationship associated with preference and belief formation to that associated with decision making, ignoring the intervening states, and relabeling the entire relationship as one of reinforcement. On the other hand, the converse is not true, at least if we wish to retain definitions of preference and belief that correspond to our everyday meanings. We would have to decompose a reinforcement relationship into two parts so that the intervening variables meet the commonplace definitions of preferences and beliefs. Undoubtedly, there are a number of behaviorist models for which this would be quite difficult if not impossible. Despite this, preference and beliefs have tended to creep back into behaviorist analysis, and this tendency has increased as psychological science has advanced. As argued earlier, the main reason for this is the usefulness of preferences and beliefs in reducing the complexity of a general actor model.

While connectionist models also introduce intervening variables, the units these variables correspond to, neurons and synapses, are hardly the basis for parsimonious representation. While it may be possible to group them into a more manageable number of aggregate units, it is not clear how this would be done without eliminating any conceivable value of connectionist models for predicting action or simply producing higher-level variables that correspond closely to
preferences and beliefs. The problem is at any rate primarily conceptual rather than empirical. One could even suggest that preferences and beliefs are simply the labels that we have chosen to attach to the most parsimonious way we have of conceptualizing individual propensities towards different behaviors.\(^{46}\) If it is the very complexity in patterns of human behavior that require the use of such concepts, the theoretical usefulness of positing preferences and beliefs will increase rather than decrease as we accumulate more detailed information on patterns of behavior.

**Rules and Norms**

Any actor model that incorporates preferences should properly be called a rationalist model (though not necessarily an optimizing one), since by definition the existence of preferences, or similar attitudinal concepts such as goals, desires, values or motivations, implies that actors will be oriented toward achieving them. For that reason, the only prominent, and perhaps the only possible, generally applicable set of assumptions about nonrational decision making that retains attitudes is one that replaces the goal-oriented actor with an actor whose decision making is largely dictated by beliefs and who is oriented towards identifying and executing of appropriate rules for action.

In social science, such a model is generally based on the premise that individuals internalize and follow norms, rules of behavior determined by their social environments. Of course, internalization is only one conceivable reason why an individual might follow a norm; the other is due to external sanctions or rewards of some sort. However, norm following based on sanctions or rewards (and hence the expected outcomes of actions) implies that individuals are driven by preferences, and can quite easily be incorporated into the conventional rational choice model. On the other hand, norm following can be seen as nonrational if individuals are assumed to follow norms without any thought about their consequences. Hence the emphasis in a nonrationalistic model of norms must be on internalization. Furthermore, a non rationalistic model would view internalization not as affective attachment (in which case it would influence preferences), but rather as a kind of cognitive assimilation.

Internalized norms are often emphasized in structural-functional theories in sociology, as well as in some organizational theories.\(^{47}\) In such theories, individual preferences play no direct role in actions and the beliefs that are most relevant are those about which rules apply in particular environmental contexts and those which identify the current context.

\(^{46}\) Or, more symmetrically, one could argue that we do not use preferences and beliefs in predicting actions because we perceive them to be genuine objects; we count them as genuine objects because we use them predicting actions.

\(^{47}\) See March and Olsen 1989, chap. 3. See also the “Model II” bureaucratic approach in Allison 1971 and the “cybernetic paradigm” found in Steinbruner 1974, chap. 3, esp. pp. 71-2. For these latter models, coercion and hence rationality might apply at some levels of a hierarchy, but the implicit assumption is that adherence to the norm at the highest level is not based on rational processes.
For a model of irrational internalized norm-following to be both general and assertive, it must contain assumptions that can predict the content of norms under any set of structural assumptions. In addition, if the assumptions do not rule out the simultaneous existence of multiple, conflicting norms, they must predict the conditions under which one norm takes precedence over another. All this, in effect, necessitates a general and assertive set of assumptions regarding the way that the environment shapes a society and how the society, without any of its members acting instrumentally, will decide upon a set of norms. Such a set of assumptions certainly appears to be quite difficult to devise.48

An alternative to such a model of passive norm absorption and structural determinism would be to posit the existence of norms about norm formation, or metanorms, that determine or not whether societies as a whole and particular groups of individuals would adopt or reject possible norm candidates. But this would simply push the problem one level back by making it necessary to account for the metanorms. Why such regress might be ended by specifying a “grand metanorm” that is innate or arises spontaneously and dictates all other norms, it is unlikely that specifying such a metanorm would provide much in the way of assertive or accurate predictions.

Perhaps because of these difficulties, theorists have been extremely hesitant to employ norms as the basis for building general decision-making models, and no prominent models based purely on norms or on nonrational decision making exist. Because of this, much of the following discussion focuses instead on the ways in which theorists have attempted to integrate norms and rational aspects of decision making into a single model.

Elster, while rejecting models of purely norm-determined action, proposes a synthetic approach that views rational choice and social norms as alternative, and both important, determinants of behavior.49 However, he does not attempt to develop this approach into a assertive model, viewing rationality and norms instead as separate, perhaps incompatible, “mechanisms” that both affect behavior, but in different ways and under different circumstances. In addition to predicting the nature of norms,50 an assertive model would also have to specify how these norms and rationality interact to determine behavior for each individual.

An alternative approach would be abandon the idea of norms as standing outside the pale of rational calculation and to incorporate as much of norm-determined action as possible into a rationalistic decision-making model.51 Internalized norms, rather than being followed without thought to consequences, could be viewed as preferences that are associated with the intrinsic pleasure of certain types of actions and/or with their consequences. Among the benefits of this approach

---

48 For a related discussion, see Goldfarb and Griffith 1991, 65-66.
49 Elster 1989a.
50 For his take on the difficulty of doing this, see Elster 1979, 125; Elster 1991
Choosing an Identity

is that it allows for calculation of tradeoffs between competing norms as well as between norms and other preferences. Goldfarb and Griffith identify this approach as “moral values as preferences,” as opposed to “norms as constraints,” i.e., norms that are followed without regard to consequences. This distinction in ways of reconciling norms with rationality is also linked to Heap’s distinction between “instrumental,” “procedural” and “expressive” rationality, the former referring to actions based upon rational pursuit of particular goals, including moral ones, and the latter two referring to actions that have intrinsic value and are nonconsequentialist. Procedural rationality is defined as action that confirms an individual’s location as a social being and hence is valuable simply for being shared. Expressive rationality refers to action that helps an individual discover his or her true preferences, and hence is valuable for providing an individual with self-respect and autonomy.

One example of an approach that incorporates norms into preferences is the well-known social psychological model of Fishbein and Ajzen. In their “theory of reasoned behavior,” an individual’s choices are based in part upon the preferences of other individuals with regards to his or her actions, multiplied by the weight that she places on complying with their preferences, and in part on his or her own independent preferences. However, this model does not attempt to specify the weight that each individual will place on the preferences of others or how an individual develops beliefs about those other preferences. Such specification requires a general, assertive model of preference formation, as does any other approach that attempts to introduce norms as a kind of preference.

Nonoptimizing Rationality Assumptions
Given the difficulty of coming up with general alternative assumptions that abandon the idea of preferences as motivating action, work on alternative decision making has focused instead on alternative ways in which actors will attempt to satisfy their preferences. In looking at such nonoptimizing rationality assumptions, I focus only on the most general models, those that apply across all possible environments, and do not attempt to analyze all of the huge variety of midrange decision-making models found in such fields as decision theory; social, cognitive, and economic psychology; philosophy of mind; ethical and moral philosophy; behavioral economics and socioeconomics; or political behavior and political psychology. Even prominent models with fairly broad applicability, such as the Michigan

52 In fact, Elster himself in an earlier work argued that “in most cases the influence is mediated through the preference structure.” See Elster 1979, 141. He also appears to return to this in later writings as well; see Elster 1991, 114-15.

53 Another category they describe is “norms as decisions rules,” i.e., rules for action that apply only under conditions of strategic uncertainty where rational optimization is impossible. See Goldfarb and Griffith 1991, 60-64.

54 See Heap 1989, chapters 7 and 8.

55 See Fishbein and Ajzen 1975 and Ajzen and Fishbein 1980.
model of voting,\textsuperscript{56} the frustration-aggression model of violence,\textsuperscript{57} and the garbage can model of decision making in organizations,\textsuperscript{58} are not examined, since they are designed to explain particular types of actions within certain specified environments. This section centers arounds four prominent types of decision-making models, drawn from various disciplines, that apply to all possible actions and environments. These are satisficing models, reinforcement models, nonexpected utility models, and normative philosophical models based on evidential logic and constrained maximization.

\textit{Bounded Rationality and Satisficing}

Many proponents of nonoptimizing rational choice models base their assumptions on the notion of bounded rationality, the idea that humans fail to optimize their preferences because of limitations in their cognitive abilities. The term is most frequently used to characterize nonoptimizing models when they are being directly compared to conventional rational choice models.\textsuperscript{59} Bounded rationality, as defined by its proponents, is more than a matter of failing to choose the best available action given one’s preferences. Even the rational optimization model allows for the possibility that actors may not choose the best available actions given their preferences if their information is incomplete. Bounded rationality assumptions take this one step further, arguing that even when individuals hold precise and complete information about all relevant environmental variables, they may fail to optimize because they do not examine all available actions fully or precisely calculate the utility of each action. This is due to their adoption of various heuristics that simplify available information and/or add additional factors that are irrelevant to optimality.

As a description of real-world behavior, bounded rationality is intuitively more appealing than either rational optimization or nonrationality. There is, furthermore, ample empirical evidence that individuals are seldom perfect optimizers, given their preferences and beliefs, but are nonetheless goal oriented.\textsuperscript{60} However, without elaboration, the assumption of bounded rationality is too accommodative of different decision mechanisms, and it does not narrow the range of possible actions that an individual may choose in any given environment. Hence any attempt to build an assertive decision-making model based on bounded rationality must specify both the nature and locations of these bounds, as well as how they affect decision-making.

The term \textit{bounded rationality} is often associated with Herbert Simon, and the most prominent general bounded rationality model

\begin{itemize}
\item \textsuperscript{56} Campbell et al. 1960.
\item \textsuperscript{57} Originating with Dollard et al. 1939. A prominent application of the model to politics is Gurr 1970.
\item \textsuperscript{58} Cohen, March and Olsen 1976 and March and Olsen, 1992.
\item \textsuperscript{59} For a survey of applications of and arguments for bounded rationality in economics, see Conlisk 1996.
\item \textsuperscript{60} Simon 1976; March 1978; Kahneman, Slovic, and Tversky 1982; Tversky and Kahneman 1986; Quattrone and Tversky 1988; and Thaler 1992.
\end{itemize}
remains the satisficing model that is associated with the early work of Simon and March on behavior in large bureaucracies.\textsuperscript{61} Since its inception as an idea, satisficing has been widely discussed in economics, political science, and sociology as an alternative to optimization and applied to a full range of decision-making environments.\textsuperscript{62} Such applications initially took hold in behavioral economics and have since become prevalent in many other economic literatures, such as the new institutional and evolutionary economics, though much of its influence is now implicit rather than explicit.\textsuperscript{63}

Satisficing involves individuals who search very selectively through large realms of possibilities in order to discover what alternatives of action are available, and what the consequences of each of these alternatives are. The search is incomplete, often inadequate, based on uncertain information and partial ignorance, and usually terminated with the discovery of satisfactory, not optimal, courses of action.\textsuperscript{64}

It views decision making as a process of problemistic search in which alternative actions are considered in sequence, consideration of an alternative may lead an individual to direct attention toward linked alternatives, and consideration makes demands on an individual’s cognitive resources. At each step of the process, individuals will decide whether to continue considering alternatives or accept the current one, basing the decision on whether the utility offered by the current alternative meets some prespecified aspiration level. Incomplete information about alternatives does not necessarily contradict the assumptions of the rational optimization model, but it is clearly suboptimal for individuals to consider only a limited number of alternatives and to choose one when available information shows that another one would provide higher utility.

Nonexamination of alternatives cannot straightforwardly be reconciled with the rational optimization model by assuming that individuals take into account the cognitive costs of calculating the expected utility of particular alternatives before deciding whether to do so. Such a model would imply that such individuals also take into account the costs of calculating cognitive costs, as well as the costs of calculating the costs of calculating costs, and so on, which would lead to infinite

\textsuperscript{61} Simon 1947; 1955; March and Simon 1958.
\textsuperscript{62} One interesting exception to this view of satisficing as bounded rationality is philosopher Slote, who argues that individuals ought to satisfice, in the name of moderation, even when they are capable of optimizing, and that satisficing is the higher form of rationality. For him, the distinction between satisficing and optimizing is the same as that between common-sense morality and utilitarianism. His, however, is a highly unorthodox view of satisficing. See Slote 1989, chapters 1-3. See also Swanton 1993.
\textsuperscript{63} See Conlisk 1996, 675. For an overview of behavioral economics, see Gilad and Kaish 1986.
\textsuperscript{64} Simon 1985, 295.
regress, making calculation impossible.\textsuperscript{65} Hence models must either ignore such costs or assume that individuals satisfice at some point.

It is clear that satisficing is a quite realistic description of individuals' decision-making processes under many contexts. For a satisficing model to be assertive, however, two types of open parameters need to be specified: the utility levels at which individuals will satisfice (their aspiration levels) and the strategies individuals use to select alternatives for consideration.\textsuperscript{66} It appears quite difficult to come up with a set of assumptions about aspiration levels and search strategies that is applicable across as wide a range of individuals and environments as the rational optimization model and is as assertive in its implications. While it is true that individuals often fail to examine all alternatives and settle for less than optimal results in a wide range of environments, there is no obvious uniformity in the aspiration levels that different individuals seek across different environments. While certain general propositions can be made about the search strategies that individuals adopt, none has been suggested within the approach that is assertive enough to make clear predictions of action without the addition of midrange assumptions specific to a particular search environment. Moreover, in highly structured environments, such as the voting booth or the ice cream shop, it is plausible that all alternatives can be examined even when all of their characteristics cannot. Here, the main problem will not be whether or not to satisfice after considering a particular alternative but determining whether or not a particular alternative is indeed satisfactory.

Rather than being the basis for a general predictive model, the assumption of satisficing seems better suited as a conceptually rich framework for midrange theory building, particularly in environments where search is indeed problemistic and where aspirations levels and search strategies can be determined inductively before being applied to making predictions about action under similar types of circumstances. While it is certainly possible (and hopeful) that future general rational choice models will be able to incorporate satisficing-type behavior, they will probably do so by deriving aspiration levels and search strategies as products of higher-level processes of preference and belief formation.

**Reinforcement**

Given that reinforcement models arise from the behaviorist approach in psychology, which explicitly avoids the discussion of mental processes, it may seem an oxymoron to refer to a reinforcement decision-making model, much less to place it under the category of rational choice, unconventional or not. Nonetheless, the assumption that behavior is oriented in a systematic but uncalculated manner toward gaining positive reinforcement is consistent with the notion of nonoptimizing rationality, even the though assumption is rarely couched in that

\textsuperscript{65} This point was first made by Winter 1964, 252

\textsuperscript{66} See discussion in Elster 1990, 42.
manner.\footnote{For an interesting comparison of rational choice and behaviorism, along with evolutionary biology, and an extended criticism of all three, see Schwartz 1986. See also the discussion of Homans’ work in the section later in this chapter on social exchange theories of value.} Hence reinforcement models can be seen as a possible source of alternative assumptions about the way that boundedly rational decision-making functions.

Perhaps the most prominent deductive model in behaviorism, Herrnstein’s “matching law,” is a model whose decision-making implications have been openly discussed and that has been compared directly to the conventional rational choice model. The basic version of the matching law states that the ratio of the amount of time spent on two different behaviors will converge to be equal to the ratio of the total amount of reinforcement each has provided.\footnote{Herrnstein 1961; 1970.} Later versions of the matching law in effect multiplied the right-hand side of the equation by the inverse ratio of average delay in reinforcement, while more explicitly separating total reinforcement into frequency and average amount.\footnote{See Chung and Herrnstein 1967. Ainslie has noted that this implies hyperbolic discounting of reinforcement; see cites in section on time preferences later in this chapter.} The matching law has been developed into a decision-making model known as “melioration,” which charts the dynamic through which individuals react to changes in the levels of utility provided by particular behaviors, with utility being explicitly equated with reinforcement.\footnote{Herrnstein and Vaughn 1980; Herrnstein 1982; Herrnstein and Prelec 1991.}

The main empirical implication of the matching/melioration model is that individuals will choose between behaviors on the basis of relative average returns rather than relative marginal returns. When certain types of behaviors exhibit declining returns in utility as a function of past frequency, this can lead to choices that contradict rational optimization, with excessive time being spent on behaviors that have provided past utility.\footnote{Herrnstein 1990 and “Behavior, Reinforcement and Utility,” in Hechter, Nadel, and Michod 1993, 137-52. For other models which also assume such utility functions, see sections on consumption-based preferences and social exchange value models later in this chapter.} Numerous experiments with nonhuman and human subjects have found this to be an accurate predictor of the way in which animals allocate their time within a set of alternative behaviors.\footnote{For a broad survey, see McCarthy and Davison 1988.}

In another reinforcement model with particular relevance for rational choice analysis, Alhadeff has adapted for economic applications the work of a number of prominent behaviorist psychologists,\footnote{Alhadeff 1982. A summary of the work can be found in 1986.} most notably Skinner.\footnote{Skinner 1953; 1969.} Alhadeff has developed a fairly complex “conflict” model of decision making oriented primarily towards explaining...
purchasing ("buy") behavior. In this model, the frequency of a behavior is related to five factors: the level of reinforcer effectiveness, the delay between a behavior and its reinforcement, the quantity of reinforcement, the quality, and the schedule according to which reinforcement occurs. The level of reinforcer effectiveness is basically the value that individuals place on a particular good, though Alhadeff avoids reference to a mental variable such as value. The schedule of reinforcement refers to whether provision of a reinforcer follows the appropriate behavior all the time (continuous) or only some of the time (intermittent). Under appropriate circumstances, the latter is seen as actually having a positive impact on the frequency of a behavior.

The level of reinforcer effectiveness is determined by past levels of reinforcement, but the manner is different for each of three main types of goods: primary (valued prior to conditioning), secondary (valued as a result of association with primary goods) and escape-avoidance (valued for reducing the level of a negatively valued good). For primary goods, the level is a function of satiation; for secondary goods, of satiation processes associated with associated primary goods and of extinction (ending of association with primary goods); for escape-avoidance goods, of changes in level of the associated negative good towards a tolerable level (a sort of negative satiation) and of extinction. Finally, for special sorts of "nonfunctional" secondary and escape-avoidance goods (those whose association with a primary good is coincidental rather than causal), the level is inassertive and actions to pursue the good are limited only by the individual's resources (e.g., income).

Whether one prefers Herrnstein’s fairly simple model of behavior or Alhadeff’s more complex but less assertive model, a number of issues must be addressed before reinforcement models can be seen as general alternatives to the conventional rational choice model. To begin with, reinforcement models are straightforwardly applicable only to environments where various subsets of a fairly small set of discrete behaviors comprise the choice sets available at any point over a long period of time since individuals are assumed to base their decisions on their past experiences with each behavior available. It is difficult to apply such assumptions to make predictions when new behaviors enter a choice set. In that sense, the environments that the reinforcement models are most appropriate for analyzing are precisely the opposite of those for which the satisficing model is most appropriate.

In complex environments, moreover, it is far from trivial to ascertain whether two behaviors available at different times will be treated as the same or different by an individual. For instance, the choice of “joining a group” can be anything from walking into a crowd to

---

75 Alhadeff 1982, chap. 3.
76 Alhadeff 1982, 144-55.
77 Alhadeff 1982, chap. 5.
78 For related discussion, see Tallman and Gray 1990, 420-22. This article also discusses a number of models of distributed decision making by sociologists that are loosely related to the behaviorist models discussed here.
form a new religion. Hence, for a general reinforcement model
doing decision making to be assertive, it must specify a general set of
criteria that determine whether an individual will conceptually equate
two behaviors. Moreover, given that two behaviors available at different
times in a dynamic environment will rarely be identical in every respect,
these criteria may need to incorporate varying degrees and multiple
dimensions of similarities in the characteristics of behaviors rather
than simple distinctions. Such sets of general criteria seem difficult
to generate, particularly if we retain the behaviorist injunction against
explicitly analyzing individual beliefs. While the problem of specifying
choice sets is common to any choice-theoretic model, it is particularly
pronounced for reinforcement models because of the need to clearly
specify the level of continuity that exists between choice sets.

Further, in many environments, the outcomes of actions may not
be obvious and often may be delayed past the onset of other actions.
In such situation, a general model needs to specify what behavior will
be reinforced, whether the most proximate one in time, the one that
had the most influence on outcomes, or some combination of the two.
Finally, accuracy requires that environmental changes must have some
impact on an individual’s choice over different behaviors, however the
behaviors are conceptualized, and that the nature of this impact be
specified. Again, it will be difficult to generate such specifications
without including beliefs in decision-making processes.

Hence, assertiveness and accuracy in making predictions outside
of highly constrained environments can be accomplished only by major
additions to existing reinforcement models, including incorporation of
cognitive assumptions. Once such a new model is proposed, it is not
clear which differences will remain between its assumptions and those
of rational optimization since the differences that currently exist are
quite specific to the environments in which the existing model has been
tested.

Nonexpected Utility
Models of nonexpected utility, as the name suggests, focus primarily
on the component of the rational optimization model that describes
decision making conditions of risk, replacing this with alternative
formal assumptions. These models draw upon theories from large
number of disciplines, primarily economics, but also psychology,
philosophy, and operations research, and have had an influence in
political science and sociology as well. One of the best-known sets
of such alternative assumptions can be found in the “prospect theory,”
developed by psychologists Kahneman and Tversky. They posit that
individuals, given a particular utility function over outcomes, act on
transformations of those functions that are loss averse, that is, risk
averse with regard to perceived gains and risk seeking with regard to
perceived losses, and that the location of the baseline from which gains
and losses are calculated is influenced by framing effects, particularly

----

79 Kahneman and Tversky 1979; Thaler 1980.
the reference point against which different outcomes are compared.\textsuperscript{80} They also posit that individuals rank different probability distributions over outcomes (lotteries) based on a valuation function that replaces the probabilities in expected utility calculations with subjective decision weights. These weights are in turn based on a nonlinear transformation of probabilities that underweights merely probable outcomes relative to certain outcomes.\textsuperscript{81}

Outside of prospect theory, several other nonexpected utility models have been proposed that posit alternative decision weights that are nonlinear transformations of probability.\textsuperscript{82} These include weighted or “alpha-\mu” choice, in which decision weights are a product of probabilities and an additional framing-influenced weighting function which can possibly be equated to “vividness” or “conceivability.”\textsuperscript{83} and rank-dependent or anticipated expected utility, in which decision weights for each alternative correspond to a function of the cumulative probability of all alternatives (in order of preference) up to and including it minus the same function of the cumulative probability of all alternatives below it.\textsuperscript{84}

Regret theory and disappointment theory, developed more or less simultaneously by Bell and by Loomes and Sugden, are two additional models that specify transformation of utility functions based upon the framing of the choice set. In regret theory, individuals are posited to minimize expected regret, which is defined as a nondecreasing function of the difference between the utility provided by the chosen action and that provided by the nonchosen action.\textsuperscript{85} For multi-action choice sets, regret is defined as a weighted sum of the two-choice regret functions for the chosen action and each nonchosen action.\textsuperscript{86} Regret theory is closely related to Fishburn’s SSB (skew-symmetric bilinear) utility theory, though the latter is based on a formal generalization of expected utility and does not attach psychological labels to its parameters.\textsuperscript{87} Both are

\textsuperscript{80} For a further discussion that extends the notion of reference points in utility functions to conditions under which beliefs are deterministic rather than probabilistic, see Tversky and Kahneman 1991.

\textsuperscript{81} In a recent update of prospect theory, the assumptions about risk-aversion have been reversed for low probabilities, and a number of other modifications have been made. See Tversky and Kahneman 1992.

\textsuperscript{82} For surveys of such models, see Fishburn 1988; Hey and Lambert 1987, chapters by Sugden and by Appleby and Starmer; Weber and Camerer 1987; Machina 1987, 90-132. A more philosophical overview can be found in Anand 1993.

\textsuperscript{83} Chew 1983; 4, 1065-92; Weber 1985.

\textsuperscript{84} Quiggen 1982; Karni and Safra 1990.

\textsuperscript{85} Loomes and Sugden 1982; Bell 1982.

\textsuperscript{86} This is somewhat different than the definition used by Savage and adopted later in this book. In a more recent paper, Loomes and Sugden present a more general form of regret theory, which allows for a wider range of functional forms. See Sugden 1987.

\textsuperscript{87} See Fishburn 1991. The concept of regret will return in next chapter as one of the bases for the model presented there, but it will be in the context of preference and belief formation rather than decision making under risk.
Choosing an Identity

related to Savage’s earlier minimax regret model of choice, though this was originally applied to choice under uncertainty rather than risk. In disappointment theory, an individual is said to maximize the expected sum of utility and disappointment, where disappointment is defined as a nondecreasing function of the difference between the utility actually provided by an action and its a priori expected utility. Both regret and disappointment-based decision making are subject to framing effects. The regret model implies that an individual’s actions may be altered by the consideration of additional alternatives, even if none of these other alternatives is ever chosen. The disappointment model implies that actions may be altered by viewing a set of actions as a single, multidimensional action, even if the consequences of different choices are unchanged.

It possible to view nonexpected utility models as models of preference and belief formation if one takes framing-induced transformations of utility functions and probabilities as reflecting an individual’s true intrinsic preferences and beliefs rather than as distortions that enter into decision-making processes. Given such a definition, individuals who maximize nonexpected utility can be seen as engaged in a form of optimization rather than as boundedly rational. However, all of the models necessarily assume the existence of some more stable underlying set of utilities and probabilities to be transformed, and it seems less confusing to limit the terms preferences and beliefs to this underlying set rather than to define two different levels of preferences and beliefs. Furthermore, such a definition accords with the notion that attitudes, even when they are changing over time, should be a set of orientations exhibiting reasonable continuity, rather than transient responses to a particular description of a choice set. This type of definition is indeed necessary if one wishes to identify an individual with a particular set of preferences and beliefs, independent of the specific choice he or she happens to be facing at the time, and to use such a set to predict an individual’s actions across a wide range of environments.

There is ample evidence that nonexpected utility models, by introducing an element of cognitive friction into the frictionless rational optimization model, help to address a wide range of important

---

88 If applied to situations of risk, Savage’s model would imply infinite risk-averseness over regret. See Savage 1951. The most prominent application of this model to social science phenomena is Ferejohn and Fiorina’s theory of voter turnout; see Ferejohn and Fiorina 1974.

89 Bell 1985; Loomes and Sugden 1986.

90 Loomes and Sugden seem to favor this view for regret, since they argue that individuals may not want to change their behavior even if they are informed that their behavior does not maximize their expected untransformed (“choiceless”) utility. See “Regret Theory,” 819-21. Tversky, on the other hand, indicated in conversation with me that he viewed the loss-aversion component of prospect theory as inconsistent with any version of optimization, and prospect theory is explicitly presented as a model of bias in decision making.

91 A similar point is made about the need to define preferences so they have sufficient persistence to explain behavior over time and across situations. Bowles 1998, 79.
empirical anomalies for expected utility maximization that have been discovered in laboratory experiments, such as the “Allais paradox,” the “Ellsberg paradox” and the preference reversal phenomenon, in which subjects make sets of choices that together are inconsistent with maximizing expected utility given any single utility function. However, in transforming utilities and/or probabilities from those in the expected utility model, they leave the magnitude of the transformation unspecified or dependent on exogenously determined weighting parameters. The extent to which such models retain the assertiveness of the conventional rational choice model will depend in large part on the extent to which efforts are made to specify in a general fashion the values these parameters will take across individuals and over time. Without such specification, nonexpected utility models become difficult to apply in a predictive rather than retrospective fashion to real world empirical phenomena.

Moreover, it is unlikely that nonexpected utility models will have the same bearing on the most prominent empirical indeterminacies and anomalies of applied conventional rational choice in sociology, political science and economics that they do for the aforementioned paradoxes. The models describe how individuals will transform given utilities and probabilistic beliefs for the purposes of decision making rather than how such preferences and beliefs will be formed. Hence they do not provide significantly greater leverage than conventional rational choice in predicting the decisions of policymakers or entrepreneurs when, given plausible structural assumptions, it is impossible to specify how such individuals generate deterministic or probabilistic beliefs of any kind about outcomes from available information. Nor do they directly address the anomalies of participation in mass collective action or voting, in which people may voluntarily engage in high-cost activity with nearly nonexistent personal material benefits. Such behavior, given plausible structural assumptions, cannot be explained by an actor model that assumes materialistic utility functions, regardless of how such preferences are transformed by endowment effects, regret or disappointment.

This is not meant to imply that such models cannot do a great deal to improve the predictive power of rational choice models. Not only do they provide a more accurate set of assumptions regarding the ways in which preferences and beliefs are turned into action than the idealized view of the rational optimization model, but they are specified in such a way that they can be used to analyze choice in all environments. Moreover, it is also quite straightforward to apply them to similar issues as the conventional rational choice model, since they adopt the

---

92 See the surveys cited at the beginning of this section for discussions of these phenomena.

93 For a critique of the alleged incoherence of psychology-based analysis in political science, see Wittman 1991.
same formal choice-theoretic framework. However, complaints that nonexpected utility models are difficult to use predictively will persist until they generate predictions of similar assertiveness to those of the conventional rational choice model in the same range of environmental settings without the need for ad hoc assumptions about key parameters in their probability and utility transformation functions.

**Evidential Logic and Constrained Maximization**

In the normative philosophical literature on rationality, the rational optimization model of decision making has to contend with numerous challenges to its privilegedness. Unlike empirically-minded proponents of bounded rationality approaches, normative theorists base their opposition to the rational optimization model not on its failure to predict action but on its failure to always prescribe actions that maximize a person’s utility. They have pointed to a number of situations in which choosing an action that is causally linked to the highest level of utility available, and thus optimal in the usual sense, leaves an individual with lower utility than he or she would have otherwise. Hence optimization is seen as contradictory to a pragmatic definition of rational choice as the action that best facilitates the achievement of one’s goals.

The debate over the normative privilegedness of the rational optimization model has not been restricted to philosophy. Some of the proponents of nonexpected utility models, particularly those who are economists, have rejected the portrayal of their models as embodying bounded rationality. However, most of their arguments are defensive, having to do with why learning effects would not cause individuals to abandon a decision-making process that does not maximize expected utility under conditions of risk, particularly in light of apparent normative anomalies for nonexpected utility. These include the “dutch book” and “money pump,” which respectively describe situations in which nonexpected utility maximizers would willingly accept lotteries they would be ensured to lose, or repeatedly pay a sum of money to trade one lottery for another whose distribution of payoffs is identical. Defenses have largely consisted of arguments that nonexpected utility maximizers would frame their choices in such a way that they would act identically to expected utility maximizers under the conditions described.

---

94 The use of prospect theory in particular is increasingly widespread in the analysis of national security policy formation. For an example of the application of prospect theory to foreign policy case studies, as well as a discussion of the methodological issues involved in making the theory deterministic, see Farnham 1994. For more general discussion of methodological issues in applying psychological models to foreign policy, see Geva and Mintz 1997.

95 For a short comment on difficulties with integrating prospect theory with rational choice to generate new predictions, see the interview with Akerlof in Swedberg 1990a, 68-69.


97 See Machina 1989. For a philosopher’s defense, see Schick 1986.
Critics of rational optimization in philosophy, however, are much more likely to claim the normative superiority of alternative models in providing utility those individuals who follow them.\textsuperscript{98} Much of this debate has taken place over normative anomalies for rational optimization under conditions of strategic interaction between individuals.

More specifically, proponents of alternative models suggest that they are superior because they are able to address cases in which optimizing individual behavior is linked to unfavorable outcomes. The most prominent include games of strategic interaction such as the finite-iteration prisoners’ dilemma,\textsuperscript{99} the chain-store paradox,\textsuperscript{100} and the centipede game.\textsuperscript{101} In each case, two or more individuals have a choice between cooperation and noncooperation in a finite sequence of interactions, and noncooperation at each stage is dominant for each individual. On the other hand, each individual would do better overall if they could jointly commit themselves to cooperating.\textsuperscript{102}

This “collective action” problem has long been recognized by social scientists as a key dysfunctionality of rational optimization, a case in which pursuit of individual optimality leads to collectively suboptimal outcomes. However, while the main response in social science has been to examine the role of social institutions in compensating for the dysfunctional aspects of rational optimization, philosophers tend to concentrate more on the ways in which alternative forms of decision making would allow individuals to gain better results regardless of the environment.

Furthermore, they extend the examination of normative anomalies to imaginary “thought experiments” such as Newcomb’s problem\textsuperscript{103} and the toxin puzzle.\textsuperscript{104} Both involve cases in which one action dominates another in its causal effects on utility but where the dominated action is known to be perfectly correlated with some past event that would

\textsuperscript{98} A broad collection of readings on this issue and on philosophical views of decision making in general can be found in Gardenfors and Sahlin 1988. Some of the most prominent articles in this debate and responses to them can be found in Campbell and Sowden 1985. Recent collections of articles by prominent names in the field include Bicchieri, Jeffrey, and Skyrms 1997; Bicchieri and Dalla Chiara 1992; and Bacharach and Hurley 1991.

\textsuperscript{99} Invented by Flood at Dresher at the RAND corporation and brought to public attention in Luce and Raiffa 1957, 100-102.

\textsuperscript{100} Selten 1978.

\textsuperscript{101} Rosenthal 1981.

\textsuperscript{102} Parfit similarly characterizes optimization as a kind of “self-effacing theory,” i.e., a theory that predicts that people who do not believe it will be better off. See Parfit 1984, chap. 1. In games involving multiple but finite numbers of interactions, the a great deal of examination has been made of the use of backward induction to iteratively remove dominated choices from consideration. See Pettit and Sugden 1989.

\textsuperscript{103} Nozick 1969, 107-33.

\textsuperscript{104} Kavka 1983.
Choosing an Identity

bring about even higher utility. \footnote{Newcomb’s problem involves a person choosing to take either one or both of two boxes when it is known that a clairvoyant benefactor will have put a prize in the first box only if she foresaw that the person would not take the second. The toxin puzzle also involves a clairvoyant benefactor, though with somewhat less benign motives, who gives a prize to a person if and only if she forsees that the person will voluntarily drink a very unpleasant toxin at some future point in time. The main difference between the two problems is that the prize in the toxin puzzle is already at hand when the decision to drink is made, so the past event in question is known by the person to have happened. While the cases are rather fantastic, similar issues can arise when the potential benefactor has only some inkling of the propensity of the person in question. This indeed is the basis for the economic model of Frank discussed later in this section.} Broad similarities along a number fronts have been noted between the prisoners’ dilemma and Newcomb’s problem, with the later being seen as something like a one-player version of the former \footnote{See Nozick 1969; Lewis 1979, 251-54; Hurley 1991; Sobel 1994, chapters 2-4.} and more generally between the different varieties of paradox. \footnote{An interesting analysis of Newcomb’s problem, the prisoner’s dilemma, the centipede game and the “liar paradox,” e.g., “this statement is a lie,” as versions of a more general “doxic paradox,” can be found in Koons 1992, chapters 1-3.}

Because of these paradoxes, philosophical critics argue that a different model of rationality is needed to make the concept meet the criterion of allowing individuals to maximize utility. Although the models that are proposed are normative ones, they can be used to address some of the perceived positive shortcomings of the rational optimization model in predicting individual choice \footnote{See cites under the shortcomings of conventional rational choice in the previous chapter.} since they specify conditions under which individuals act cooperatively, even when doing so is inconsistent with rational optimization. While a large number of alternative models have been proposed, two types have attracted the most attention, the evidential logic model, also known as the conditional expected utility model, and the constrained maximization model, also known as the resolute choice model. As the multiple names suggest, each type of model has been suggested by more than one theorist, often with certain differences in specification. However, the similarities within each category are greater than the differences. Both types of models have been compared extensively to the standard assumption of rational optimization and are framed in choice-theoretic terms.

According to the evidential logic model of decision making, individuals optimize over actions but may take their own actions to be sources of information about the environment. \footnote{Such a model was first formalized by Jeffrey. See in Jeffrey 1992, chap. 13. It is presented in full detail in Jeffrey [1967] 1983. An early mention of this type of model can be found in Nozick’s 1963 PhD dissertation, which remained unpublished until 1990. See Nozick 1990, 232.} Under risk, they use subjective probabilities over outcomes to calculate expected utility, but the probabilities for each action are conditional on the choice of action (hence the other major term for this type of model, conditional
expected utility). This in turn implies that expected utilities over actions are not necessarily based upon the causal effects of actions.\footnote{For discussions, see Jeffrey [1967] 1983, 83-85, 157-58; Eells 1982, 79-82; Sobel 1994, 141-45; Nozick 1993, 41-50; Grafstein 1992, 103-8.}
The most frequently used historical example of apparent evidential logic is that of the Calvinists during the Industrial Revolution, who (at least according to one common interpretation) believed in divine predestination but nevertheless vigorously pursued worldly success in order to prove to themselves that they were among the elect.\footnote{Weber [1904] 1958; Jeffrey [1967] 1983, 25; Eells 1982, 91; Nozick 1993, 46.}

Some of the supporters of the evidential model favor it primarily for its formal simplicity and flexibility\footnote{Jeffrey’s formalization, building on work by Bolker, is quite different from those in the causal expected utility models of Ramsey, Von Neumann and Morgenstern, and Savage. The usual ontological distinctions are not made between between states, outcomes, and actions. Each is viewed as a kind of proposition that can have positive or negative desirability (utility) attached to it and can provide information that is used to update the subjective probability of other propositions. The only unique characteristic of actions is that individuals have direct control over the truth value of actions in their choice set. See Jeffrey [1967] 1983, chapters 4-5; Bolker 1967. See also discussion in Eells, 146-49.} and note that, given appropriate restrictions on beliefs, its predictions are no different from those of the rational optimization model.\footnote{Eells 1981, and Jeffrey 1981; Eells 1982, chapters 6-7; Jeffrey [1967] 1983, chap. 1, sec. 7.} However, other proponents eschew such restrictions and view the resulting differences in predictions between the evidential model and the rational optimization model as the former’s major virtue. Given the assumption that each player believes that other players are essentially similar to him or her, face similar choices, and hence will always make the same choices that he or she does, evidential logic will make cooperation rational in the finite iterated prisoners’ dilemma game, even if each player’s actions has no causal effect on the actions of the other.\footnote{Nozick 1993, 50-59. Grafstein 1992, 112-19.}

Furthermore, evidential logic also makes it rational to choose an option in Newcomb’s problem that is causally suboptimal but is perfectly correlated with a favorable outcome.\footnote{Nozick 1993, 41-50, Grafstein 1992, 119-22.}

The evidential model, despite its origins in normative theory, has been applied to explain empirical phenomena, primarily by the political scientist Grafstein. In one theory, he shows how self-interested voters following evidential logic might vote, even if the benefits of voting are lower than the costs, because their decision to vote will determine their beliefs about whether or not others will vote as well.\footnote{Grafstein 1991. See also Grafstein 1992, chap. 7. This theory in essence formalizes the notion of “if I don’t vote, no one will.”}

In another theory, he shows how a perceived similarity with members of a group can cause an individual to act more cooperatively towards them, independent of any causal implications, since his or her own cooperation will be seen as providing information on the future behavior.
of the other group members towards him or her. This is then used to explain how the phenomenon of ethnic group mobilization might transcend conventional rational choice without necessarily involving altruism.\textsuperscript{117}

In the constrained maximization model of decision making, individuals are assumed to optimize over rules for actions rather than actions themselves. Once a set of rules is chosen, it will always be followed. This in effect implies that individuals can commit themselves to following a particular strategy even when sticking to the strategy will be suboptimal for them at some later point in time. This model was brought into prominence by Gauthier in his contractarian theory of morality.\textsuperscript{118} In Gauthier’s formulation, constrained maximizers are contrasted with straightforward maximers, i.e., conventional optimizers. Gauthier argues that individuals who are rationally bargaining in a state of nature and are able to choose their own dispositions will choose to become constrained maximizers committed to cooperation rather than straightforward maximizers. Furthermore, individuals who are “translucent” in the sense that their dispositions are partially visible to others would benefit from unilaterally selecting a disposition toward constrained maximization, since this will make them a more attractive party to any such agreement.\textsuperscript{119}

A similar, and somewhat more formally specified, model was proposed by McClennen.\textsuperscript{120} Like the constrained maximization model, this resolute choice model assumes that individuals can commit themselves to rules of actions and ensure their own adherence to these rules. The main difference between resolute choice and constrained maximization is less in the specification of the models themselves than in the way the authors characterize the relationship between the models and the concept of rationality. Resolute choice is portrayed as the conscious abandonment of local (act-specific) rationality for the sake of global (overall outcome) rationality, therefore implying that contradictory decision-making processes may nonetheless both be rational. Constrained maximization, on the other hand, is portrayed as a higher form of rationality than optimization in the sense that it reflects individuals having control over their dispositions, which in turn have control over their actions.\textsuperscript{121}

While the constrained maximization/resolute choice model has remained largely within normative theory, it is similar in content to certain models proposed in the social sciences. One related model

\textsuperscript{117} Grafstein 1995.


\textsuperscript{119} Gauthier 1986, 174-77. See also discussion in Parfit 1984, 7.

\textsuperscript{120} Originally put forward in McClennen 1985, and elaborated in McClennen 1990.

in the economics is that of Frank, who uses the term commitment model to refer to emotional dispositions that lead to suboptimal choices but nonetheless may allow an individual whose disposition is partially visible to ultimately do better in attaining utility. Unlike either Gauthier or McClennen, he views the commitment model as the adoption of a local utility function that conflicts with the one an individual has at the global level, i.e., "tastes" versus "objectives," rather than an alternative way of maximizing some unified utility function, though he does not assume that individuals can freely choose the disposition they prefer.

Similar work has been done by sociologist Vanberg, who coins the term rule-following behavior to describe the tendency to follow particular rules of behavior even when they are not optimal for each action. He argues that while such behavior is nonoptimizing it can be justified by computational costs and the possibility of misidentifying the situation as well as by reputational effects. Hence, he comes closer than any of the other authors to characterizing his model as simply embodying a more nuanced version of optimization of actions over outcomes.

While the assumptions of evidential logic and constrained maximization are very different, there are significant parallels between them, aside from simply the anomalies they are aimed at addressing. Both argue that there is some characteristic of decision-making processes, other than the causal consequences of decisions themselves, that enters into utility calculations. Furthermore, both models extend the sphere of choice, at least indirectly, beyond the choice of action. By choosing an action based on evidential logic, an individual is indirectly choosing to have certain beliefs. By choosing a disposition, an individual may in part be choosing the kind of utility function that he or she will have. Indeed, both models open the door to ways in which assumptions about preference and/or belief formation can be integrated with assumptions about decision making, and how the one-way relationship of the conventional rational choice model between preferences, beliefs, and actions can be turned into a more complex and mutual one. In doing so, they provide models that can directly address many of the empirical anomalies of the rational optimization model as well as the normative ones. These include the key anomalies that seem to involve actors sacrificing utility without compensation, such as voter turnout and costly collective action.

However, this strength of the normative philosophical models in bringing about potentially greater accuracy generates the weakness, at least in the empirical sense, of lowering their assertiveness. Neither the evidential logic nor the constrained maximization model assertively specifies the limits on preferences or beliefs. Of course, assertive prediction may be seen as irrelevant to normative theorizing since such theorizing can be seen as applying to idealized beings and worlds rather

---

123 Vanberg 1994, chapters 1-4.
Choosing an Identity

than concrete reality. However, to the extent that fulfilling normative criteria of rationality depends on generating clear prescriptions for action, assertiveness is to the benefit of normative models as well as empirical ones.

While the evidential model assumes that individuals take their own actions into account in updating their beliefs, the model does not provide general assumptions about the prior beliefs that they are updating and therefore about the direction or magnitude of the effects of their actions on their beliefs. In particular, it is necessary to specify the underlying parameter with which individuals identify their own actions as well as their beliefs about the distribution of such a parameter if it is random. Furthermore, there exists no obvious criterion of superficial plausibility that can be applied to such assumptions, whereas such a criterion does exist for assumptions about the types of prior beliefs that individuals might have about the environment. Part of the reason for this is that most common-sense norms of plausibility are governed by an implicit causal logic. This makes it hard to limit the kinds of ad hoc assumptions that can be introduced into the evidential model when it is applied to generating theories for particular empirical settings.

Specifying general assumptions in an assertive manner requires some way of delineating the ways in which individuals will draw conclusions from their own actions for future actions. This includes assumptions about criteria through which individuals draw evidential connections between their own actions and those of others as well as those through which they draw connections between actions which take place at different times under less than identical circumstances. Nor, as shown in the Calvinist case, is evidential logic limited to the implications of their actions for future actions; it can also be applied to implications for the past, current, or future state of the environment, and some constraint has to be put upon these kinds of conclusions as well.

A similar need to specify limits needs to be made for the constrained maximization model if it is to be applied in an assertive manner to a wide range of empirical phenomena. If an individual can at least in part choose his or her own disposition, what determines the range of dispositions that can be chosen and to what level of specificity do strategies need to be specified? Furthermore, if dispositional choice occurs at more than one one point in time, when are such choices possible and what are the limits at each point? Are commitments to disposition permanent, and if not, what determines the point at which they can be unilaterally abandoned or collectively renegotiated? Finally, if individuals are translucent with regard to their dispositions, to what extent does such translucence exist? Questions of accuracy also exist, since it seems intuitive that even translucent individuals would do best not by choosing a permanent disposition toward constrained maximization, but rather by choosing a temporary, though genuine, disposition that will be abandoned once other parties have made their
commitments. This is not a moot question empirically since it relates to the practice of self-deception, and the possibility that such a practice could be beneficial to an individual, allowing him or her to create a credible but false image of commitment.

**Summing up Nonoptimizing Models**

Overall, many nonoptimizing rational choice models have the virtue of providing greater accuracy of prediction than a rational optimization-based model under important sets of circumstances, and the ones examined here also retain the general applicability of the model. However, they do so at a certain expense of lost assertiveness and parsimony. The former in particular makes it difficult to use such models to generate predictions in specific environments without the addition of a significant number of ad hoc assumptions, not only about structure but about the actors as well.

This problem obviously does not negate the usefulness of such models. Even when their predictions are inassertive, they often provide more realistic descriptions of the ways in which abstract variables like preferences, beliefs, and actions link themselves together and more adequate explanations of the concrete process by which individuals make decisions. Nor should nonoptimizing rational choice models should be seen as inconsistent with new models of preference and belief formation. Rather they are complementary, since each type of model addresses different empirical problems in the conventional rational choice model. Further refinement of nonoptimizing models may specify them in a way that allows assertiveness to be improved without sacrificing generality, accuracy, or parsimony. However, until this is accomplished, they cannot be viewed as filling the rational optimization model’s role in theory-building, nor solving its major empirical problems.

**Preference and Belief Assumptions**

Even if one chooses to replace the rational optimization model with alternative decision-making assumptions, the previous discussion indicates that addressing the indeterminacies and anomalies that arise in conventional rational choice applications to social, political and economic phenomena also requires new assumptions about preferences and beliefs. To a large extent, the contrast between the conventional rational choice view of preferences and beliefs and alternative views corresponds to the frequently made distinction between “materialist/realist” or “objective” views of human nature and “idealist” or “subjective” views. One important thing to note about this distinction is that it is not between competing models of action

---

124 See Smith 1991b.
125 For a concise discussion of how self-deception can be adaptive, see Wright 1994, chap. 13. On its employment in male-female relationships, see chapter 3, p. 62. For a collection of writings on this topic, see Lockard and Paulus 1988.
126 For further discussions of this contrast, see Pye 1965, chap. 5; Alexander 1990. Debates over this issue have become particularly prominent in the past few years in international relations. See Kegley 1993.
Choosing an Identity

but rather between a fairly unified and narrow model and a wider viewpoint that encompasses numerous models. Furthermore, idealism does not rule out a role for materialistic preferences, and subjectivism does not rule out a role for objective beliefs, however defined. Hence, the main task for opponents of the conventional rational choice model who wish to match its main predictive strengths is to specify the form of idealism and subjectivity that will characterize a particular individual at a particular point in time.

As was noted in chapter 1, alternative assumptions about preferences and beliefs can be found, though covertly, within purportedly conventional rational choice theories. In particular, the number of theories that achieve tractability or assertiveness by introducing ad hoc assumptions about beliefs, unexplained in terms of observable information, is quite large. Numerous incomplete information models can be put in this category since such models often rely on ad hoc assumptions about actors’ prior probabilistic beliefs, without making it clear how or where such beliefs were acquired from observation.

For preferences, there is a large and growing body of literature that is based on an overt and self-conscious effort to investigate sources of utility beyond those specified in the conventional model. One prominent theorist in this area is Becker, who, despite his reputation as a defender of conventional rational choice, has inserted such assumptions as a “taste for discrimination” into his theories of social interaction between races and altruism between parents and children into his theories of the family. In his wake have followed theorists who have extended this type of work to other, primarily political, arenas. These include Johnson, who inserted a taste for nationalism into the utility functions of Third World leaders; Tullock, who posits an “entertainment” utility from participation in revolution; and Wintrobe, who differentiates between rulers who maximize wealth and those who maximize power. In the economic arena, Lane has distinguished between utility gained from money and utility gained from work satisfaction.

There is no doubt that midrange theories based on such broadened notions of utility introduce greater verisimilitude to analysis of particular phenomena than those using conventional assumptions. However, each set of assumptions is designed to analyze individual behavior within a specific context rather than across all contexts. Hence none offers a direct challenge to conventional rational choice as a general model. Leaving the general specification of preferences open and simply allowing theorists to insert arbitrary assumptions into midrange theories would make the rational choice model as a whole

---

127 In additional to those mentioned here, additional lists of such theories can be found in Mansbridge 1990b, 256-57; Etzioni 1988a, 26-27; .
129 Johnson 1965; Tullock 1971; Wintrobe 1990.
130 Lane 1991, chapters 13 and 22.
Alternatives to Conventional Rational Choice: A Survey

Tautological, since any observed variation in behaviors could then be attributed after the fact to differences in preferences.

Because of their manipulability, ad hoc assumptions about preferences have been criticized as "empty" and "not an explanation of behavior but merely a ghost that gets blamed for observed events." Moreover, altering assumptions about preference on a case-by-case basis without an underlying justification leaves theories open to criticisms of inconsistency and bias. For instance, England notes that economists tend to adopt contradictory assumptions about the preferences of individuals, depending on whether they are modeling market behavior and behavior within a household, and that they make little attempt to justify these differences. McLean points out that rational choice theorists often make quite contradictory assumptions about the preferences of economic actors and those about political actors.

These criticisms imply that more effort needs to be made to ground the assumptions about preferences and beliefs found in midrange theories in some general set of assumptions, hence reducing their manipulability, without resorting to conventional assumptions. This puts a focus on the social science literatures that specialize in studying the formation of preferences and beliefs.

As Almond says of rational choice theorists, "As the blank tile in scrabble can take on the value of any letter, so the rational choice assumption, they seem to be assuring us, can take on the value of any utility imputed to it. Given this viewpoint it is difficult to justify their neglect of the social science literatures that display the variety of values, preferences and goals in time and space." Or, as rational choice theorist McLean puts it, "The bus has brought you so far down this road... but it stops here. To go further, you must get on the connecting bus run by the psychologist (or anthropologist, or historian) over there."

Akerlof is the most notable among the rational choice theorists who have adapted insights about preference and belief formation from various empirical social science literatures to generate midrange theories. He incorporates well-established psychological, sociological, and anthropological concepts (e.g., reference groups, cognitive dissonance and social identification) into his assumptions about individual preference structures. These types of theories are much less vulnerable to the accusation that their assumptions have been manipulated since they

---

131 Samuelson quote in Okun and Perry 1973, vol. 1, 10. See Hirschman and Wildavsky articles cited in previous chapter and Etzioni 1988a, 28-31 for further discussion and citations to other similar arguments.
134 Almond 1990a, 135.
are justified by reference to empirically supported sets of propositions from non-rational-choice literatures. Furthermore, the use of insights from empirical literatures increases the likelihood that such theories actually portray the processes by which choices are made, rather than coincidently mimicking their outcomes, and hence makes the theories more likely to be generalizable outside the context for which they were designed.

Nonetheless, building a general and reasonably parsimonious model of preference and/or belief formation from existing social science literatures presents some major challenges. The relevant social sciences are far from unified on the issue of preference and belief formation, and assertive models tend to focus on specific environmental contexts rather than general explanations. Hence, the choice of appropriate assumptions for a general model is far from obvious. This may explain why those in other social science disciplines have seemed hesitant to pick up the ball where economists have dropped it. As sociologist Hechter warns of the task of developing general models of preference formation: “Because of its difficulty, the economists—with their customary generosity—have snarkingly bequeathed it to the sociologists. Not for nothing do they believe in rational choice! We should certainly look this gift horse carefully in the mouth.”

Despite these obstacles, there is strong support for an alternative general model of preference formation, though there is as of yet little agreement on the precise form such a model will take. While criticisms of the conventional model of belief formation are also rife, there are fewer calls for a general model of belief formation. One possible reason may be that the incorporation of incomplete information has already provided the conventional rational choice model with a somewhat dynamic view of beliefs. Furthermore, belief formation outside of information has been used in the social sciences more to retrospectively identify cognitive biases and limitations than to make predictions, which in turn may make a incorporating such fact into a general model seem far-fetched.

As in the previous section, I do not attempt to survey all social science theoretical literatures that have a bearing on the issue at hand, limiting myself to the more manageable tasks of examining assumptions and models that are couched in choice-theoretic form and are regularly compared to the assumptions and models of conventional rational choice. At any rate, while there are numerous alternative sets of assumptions of this sort about preferences in various social science

---

137 See Hechter 1986a, 273.


139 See cites in section on the shortcomings of conventional rational choice in the previous chapter.
In the literatures, there is really no comparable set of alternative assumptions about beliefs.

Different alternative preference assumptions and the models they are drawn from will be discussed in the following order: I first discuss various types of models that essentially assume static preferences but different ones from those assumed by conventional rational choice. This begins with a discussion of alternative formalizations of utility based upon noncardinal, noncompensatory preferences, in which utility is calculated through fairly complex algorithms rather than simple mathematical functions. I then discuss various alternative substantive utility functions, focusing on alternatives maximands to personal material welfare proposed by two different literatures, evolutionary biology and moral philosophy, then moving on to discussing various multidimensional utility functions proposed in a variety of different literatures. Following that, I move onto dynamic models of preference formation. These include models of time-based preference change triggered by nongeometric discounting of utility and models of consumption-based preferences change triggered by addiction, satiation, and rates of change in consumption. The final set of dynamic models examined are found in the sociological literature on social exchange. Some of the literatures use terminology such as desire or value rather than preference per se. However, these concepts are generally viewed as being synonymous with preferences in the literatures discussed, and they are assigned a similar role in the determination of action.

Perhaps the simplest way to modify the conventional assumption that individuals value personal material welfare is to replace this with an alternative maximand that is also isomorphic across individuals, and there are two prominent theoretical literatures from which such a maximand might tentatively be drawn. Both in turn have drawn influences directly from the choice-theoretical framework of rational choice, and their models can thus easily be adapted to it.

**Biological Models of Inclusive Fitness**

The first is found within evolutionary biology, where theorists in the subdiscipline once known as sociobiology and now known as evolutionary psychology have long used the concept of inclusive fitness to represent the maximand that is pursued by all animals, whether human or not, and to form the basis for a general model that can explain all human action. This model comes with clearly specified maximand that is very different from the conventional rational choice maximand of personal material welfare. It is the concept of

---

140 The change of terminology reflects evolutionary psychology’s greater emphasis on the mediating effect of conscious thought processes between genes and behavior. Another, perhaps, may be a pragmatic attempt among evolutionary psychologists to distance themselves from the social Darwinist label that has long been attached to sociobiological thinking. For short discussions of the politics of terminology in the field, see Horgan 1995, 176.

141 In fact, its most prominent book more or less proclaims that sociobiology will eventually subsume the other social sciences. See Wilson 1975, 3, 547.
inclusive fitness,\textsuperscript{142} sometimes also referred to as kin selection, which is generally defined as the frequency with which one’s genes reappear throughout all future generations. In the evolutionary long run, it is argued, selection pressures ensure that organisms that fail to maximize inclusive fitness will tend to disappear, hence all organisms will devote themselves to pursuing it. Even though there is no rationality assumption in evolutionary psychology, proponents argue that selection in favor of inclusive fitness maximization will lead individuals to act as if they engaging in optimizing calculations, regardless of whether or not such calculations actually occur.\textsuperscript{143} One difficulty that proponents of the inclusive fitness model have in making their model compatible with rational optimization or even bounded rationality is that rational pursuit of inclusive fitness will require individuals to quickly adjust their behavioral patterns in light of changing circumstances rather than waiting for natural selection to generate such changes innately. The primary way in which that evolutionary psychologists reconcile such rapid adaptation with their assumption that behaviors are based upon natural selection is by assuming that rationality itself is a result of evolutionary processes.\textsuperscript{144} This assumption protects the model against the argument that it is internally inconsistent due to its failure to apply evolutionary assumptions to all aspects of human behavior. However, even if it superficially increases the range of phenomena addressed by natural selection, it is also a major concession regarding the predictive power of any theory of human behavior based upon natural selection. Even if rationality is an evolved characteristic, its ability to override just about any genetically programmed behavior means that the link between genotype and behavioral phenotype will become extremely loose, which then brings into question why it cannot override any purported genetically programmed urge toward inclusive fitness. Furthermore, if humans are credited with genuine rationality rather than instincts programmed to approximate rationality, then they must be aware of the utility function they are seeking to maximize. This is rather anomalous, since presumably very few people in modern societies would perceive long-term genetic propagation as their main goal in life.

Regardless of what position one takes on the issue of the compatibility between rational decision making and evolutionary determinism, it is questionable whether inclusive fitness can provide much leverage in addressing the predictive shortcomings of conventional rational choice. While it is clear that maximization of inclusive fitness can lead to significantly different choices than maximization of personal

\textsuperscript{142} Originally coined by Hamilton 1964. An explicit equation of inclusive fitness to utility can be found in Dawkins 1994.

\textsuperscript{143} Note the similarity to “as if” functional/evolutionary explanations for self-interest maximization discussed in chapter 1. See also discussion in Schwartz 1986, 195-201.

\textsuperscript{144} This is an argument that has intrigued philosophers of mind and cultural anthropologists as well as rational choice theorists and evolutionary biologists. See for instance essays in Dupre 1987 for a collection heavy with philosophers and Barkow, Cosmides, and Tooby 1992 for one primarily by social scientists.
material welfare, these differences are most obvious in choice situations that directly involve issues of kinship and reproduction, whereas the differences are less clear for the majority of other social, political and economic issues. Even such communal issues as ethnic conflict are difficult to address from a kinship point of view, since members of larger ethnic groups will share so little common genetic material that it is unclear what effect in-group/out-group differences will have on collective action.

In the analysis of most social settings, the assumption that individuals will rationally optimize inclusive fitness is radically indecisive. Because it focuses on genetic propagation over several generations, the assumption can accommodate evidence of actions (such as voluntary childlessness) that might seem anomalous for narrower definitions of fitness. However, it also makes the assumption very difficult to test. Because it focuses on genetic propagation over several generations, inclusive fitness, unlike personal material welfare or number of children, is a maximand whose final value cannot be measured directly. This value indeed will never be known in the lifetime of either an individual or any social scientist studying him or her, and will depend on influences that neither is in any position to anticipate. This lack of information allows a theorist to always find plausible ad hoc structural assumptions for why a behavior that seems maladaptive for short-term genetic propagation may be adaptive for very long-term genetic propagation. Such explanations, however, bring back accusations that such theories cannot be used predictively and are simply an attempt to tag with an adaptive label any behavior that has already occurred.\footnote{Famously dubbed “just-so stories” in a polemic by the Sociobiology Study Group of Science for the People, a group of anti-sociobiology scholars. See Sociobiology Study Group of Science for the People 1976, 280-90, 287. See also Gould and Lewontin 1979.}

**Philosophical Models of Rational Preferences**

A very different alternative model of preference is found in moral philosophy and philosophy of mind, where theorists have long analyzed preferences from a normative point of view, much in the same way that other theorists have examined decision-making normatively. And like the normative models of decision-making, normative models of preference can be said to have empirical implications to the extent that they represent ideals that individuals may aspire to fulfill. Preferences are analyzed according to two major criteria, rationality and morality. Naturally, the emphasis of the discussion in this section is on the former, though a clear distinction is not always made within the literature.\footnote{Normative analysis of preferences is extremely rare in the social sciences, but a few authors have commented normatively on the social effects of teaching empirical models based on conventional rational choice assumptions. See Steiner 1990 and Frank, Gilovich, and Regan 1993.}

Philosophers who write about rationality can be divided into two schools of thought regarding preferences: “instrumentalists” argue that any coherent set of preferences is compatible with rationality, while “objectivists” argue that certain types of preferences are intrinsically...
Choosing an Identity

irrational. While there is no consensus among objectivists as to the exact composition of rational desires, discussion tends to center around the closely related notions of rational desire or real interest, as judged by some standard relating to an individual’s welfare needs and/or role in society.

Gert identifies irrational desires as those that are harmful to the individual, including such things as desire for death, pain, disablement, deprivation of freedom, or deprivation of pleasure. In a variation of this argument, Parfit argues that irrational desires are those that cannot be linked to something of objective worth, such as pleasure or the avoidance of pain. On the other hand, Hollis locates real interests not so much in physical welfare but in an individual’s “identity as a social person.” This means in part that a rational person will seek to fulfill the normative expectations embodied by his or her role in society or in an organization. Combining these two views, Rescher describes real interests as those “meeting the needs that people universally have in common—health, normal functioning of body and mind, adequate resources, human companionship and affection and so on.” They are also, however, “a matter of the particular role one plays: co-operative children are in the interests of a parent, customer loyalty in those of a shopkeeper.”

Related definitions are emphasized in recent theories of justice in order to specify the nature of relevant goods to be distributed within a society. Particular emphasis is put on goods that any rational individual would want; these in turn are thought to be closely linked to a person’s physical and social welfare. Such maximands include Rawls’ concept of “primary social goods,” which include “rights and liberties, powers and opportunities. income and wealth,” and Sen’s concept of “capabilities.”

147 For a discussion of instrumentalism vs. objectivism, see Audi 1995, overview chap., 3, 25-26. A similarly-used pair of terms is “internalist” vs. “externalist,” which relate more generally to the issue of whether or not attitudes can be fully characterized without reference to the environment. See Williams 1979, 101-13. Still yet another common pair is “Humean” vs. “Kantian,” after Hume’s A Treatise of Human Nature and Kant, primarily his The Fundamental Principles of the Metaphysics of Ethics. However, Kantianism can also be understood more narrowly as a specific criterion based on Kant’s categorical imperative. See Sugden 1991, 753-56; Gauthier 1986, 6-8; Hollis 1987, 74-91.


149 Gert 1988, 28-34.

150 Parfit 1984, 120-27.

151 Hollis 1987, 91. See also Hollis 1996, chap. 2; 1983.


153 Rescher 1988, 100. See also Rescher 1993, chap. 3. Rescher actually divides rationality into three components: evaluative, cognitive, and practical, relating to preferences, beliefs, and actions, respectively. See above and also Rescher 1992, chap. 1.


Finally, there are normative models that emphasize the desires that one would adopt if one were to deliberate at length in an informed manner and were free from the grip of momentary passions. The most prominent recent version is that of Brandt, who refers to the process by which such desires would arise as “cognitive psychotherapy.” While this last group of models does not clearly distinguish between intrinsic and extrinsic desires, they do have implications for the former, as exemplified by the fact that Brandt deems irrational any desire that is purely a result of cultural transmission or exaggerated reaction to early deprivation experiences, and considers rational the desire for “native” (e.g., innate physical) pleasure or avoidance of pain.

The main difficulties in adapting notions like rational desires or real interests to an empirical model not only include the need to select from or reconcile the varying definitions proposed by different theorists, but the need to unpack the various components of physical health, freedom, and social identity; to specify the various weights an individual will place on each; and to specify ways of measuring their levels for each individual. Each of these steps is necessary if the resulting model is to be decisive. Such an enterprise clearly requires a fundamental shift in the way in which philosophers have approached the concept of desires. And, while such steps may be less important for normative than for positive models, it would seem helpful nonetheless to prescribe in a more decisive manner the content of rational preferences in order to make clearer recommendations for rational action.

While both evolutionary biology and moral philosophy arguably provide a richer model of human motivation than is provided by the arid wealth-maximizing notion of the conventional rational choice model, they both do so at a significant cost to decisiveness. The maximand of inclusive fitness generates indecisiveness because it is impossible to measure directly, and its expected value can vary too greatly within the space of plausible structural assumptions, even when beliefs are restricted to those that follow from available information. The maximand of rational desire or real interest, regardless of the version one adopts, generates indecisiveness because it is extremely abstract and is a portmanteau concept containing multiple, more concrete concepts that need to be weighted and measured separately.

Overall, while other alternative maximands might also provide better predictions than wealth in some environments, they would also suffer from similar or greater problems of indecisiveness and/or inaccuracy when used as the basis for a general preference model.

---

156 One early expression of such view of is Sidgwick 1907, 111-12. Parfit calls this view of preferences a “deliberative theory” of rationality, while one based on some fixed notion of real interest is called a “critical present-aim theory” and instrumentalism is called a “present-aim theory.” See Reasons and Persons, 117-20. Gert refers to deliberative theories as “cool moment” theories. See Gert 1988, 27-28.

157 Brandt 1979, 113, 126-27.


159 Brandt 1979, 130-32.
Highly concrete ones such as length of life or number of children would suffer from narrowness and consequent inaccuracy in making predictions, and it is difficult to think of one that would generate fewer anomalies that the wealth maximand. Highly abstract ones such as “happiness,” “goodness” and “satisfaction” might be more accurate but only at the cost of decisiveness. Precisely because of their abstractness, they would suffer from a great lack of consensus on how to measure them (if this is possible at all) and on admissible structural assumptions about the huge number of environmental factors that could conceivably influence their expected future level. If abstractness were not a problem, one could just as well posit that “utility” itself is the maximand without specifying how it is determined.

**Substantive Multidimensional Models of Utility**

Another alternative to a single alternative maximand is to posit utility functions that are multidimensional. Several authors have suggested two-dimensional utility functions that have general applicability across environments. These include Scitovsky’s contrast between comfort and pleasure; Lindenberg’s between physical well-being and social approval; Offer’s distinction between interests and social “approbation”; and Etzioni’s between self-interest and moral commitments. Hirsch, Boudon, and Frank have each suggested utility functions comprising personal welfare and relative social status, while the most commonly proposed two-dimensional utility function of all, which is found in models of altruism, comprises personal and group welfare.

Although to my knowledge there has been no attempt in rational choice models to identify more than two dimensions for a generally applicable utility function, such a utility function might be drawn from psychological works on motivation. Prominent candidates include Freud’s discussion of the two types of id instincts (eros and thanatos) and the superego’s moral values; Maslow’s hierarchy of needs ranging from physiological to safety to belongingness to esteem to self-actualization; and McClelland and Atkinson’s analysis of the needs for achievement, affiliation and power. While the issue has been less central to their concerns, sociologists have also been active in

160 Scitovsky 1976, chap. 4.
163 Etzioni 1988a, part I.
164 Hirsch 1976; Boudon 1977, especially chapter 5; Frank 1985.
166 See “The Interpretation of Dreams” and “Beyond the Pleasure Principle” in Freud 1950.
168 McClelland et al. 1953.
Whatever their form, multidimensional utility functions raise questions not only about the nature of the variables contained within them but about the relative weights to put on each variable. Any multidimensional utility function with weights that are constant across individuals and over time would have a substantial amount of arbitrariness to it, and no one to my knowledge has suggested one. On the other hand, specifying values for such weights in a way that accommodates variation between individuals and over time would require a general model of preference formation. Some candidates are discussed later in this chapter.

Formal Multidimensional Models of Utility

The multidimensional models of utility just described retain a fairly simple and usually linear mathematical form or do not formally specify their functional form. However, considerable work has been done in economic psychology and decision theory to develop different ways of specifying the way in which multiple maximands are aggregated to determine overall utility. Rather than specifying the substantive concepts linked to particular variables in a utility function, these models focus on the nature of the function itself. Most of these models avoid the need to specify utility weights by abandoning mathematical formulas altogether and employing algorithms that examine each utility variable in sequence or parallel to eliminate choice alternatives from consideration. Such models are referred to as noncompensatory in the sense that each stage of the weeding out process involves a single utility variable, and hence higher values of one variable do not compensate for lower values of another at any single stage.

It could be argued that such models represent models of decision-making rather than preference, but I chose to put them in this section because they are not couched as ways of optimally or boundedly maximizing any underlying preference ordering, but rather as a way of determining an ordering over certain outcomes known to be associated with each choice. Furthermore, it is legitimate to call this a preference ordering, since it represents the relative desirability of outcomes. Since the preference orderings are noncardinal and incomplete (suboptimal alternatives are not necessarily compared to one another), prediction of choice under conditions of risk becomes problematic. Nonetheless, suitable modifications of these techniques might be adaptable to more traditional notions of utility as well.

There are two major methods for determining noncompensatory preference orderings, threshold rules and lexicographical preferences. Threshold rules include conjunctive rules, which select alternatives that can meet designated thresholds for all of a set of specified variables,

\[169\] A discussion and synthesis can be found in Turner 1987.


\[171\] For surveys noncompensatory models, see Bettman 1979 and Earl 1983, chap. 4.
and disjunctive rules, which select alternatives that can meet the threshold for any variable in such a set. Although conjunctive and disjunctive rules eliminate the need to specify utility weights for each variable, they substitute the need to specify threshold levels for each variable. Furthermore, they do not always isolate a single optimal outcome (which necessitates tie-breaking rules) or may rule out all outcomes (which necessitates rules for selecting default alternatives).

This problem of indecisiveness is reduced by more complex methods that involve testing thresholds in sequence based on a ranking of variables or even by random selection, but these methods introduce additional ranking variables whose values remain unspecified.

For lexicographical preferences, variables are examined according to a specified precedence ranking, and any alternative that provides the highest expected value for the currently examined variable is selected without further consideration. Specifying a precedence ranking is arguably a less arbitrary matter than specifying utility weights or threshold levels since it requires assigning an ordinal rank rather than a precise number to each variable. Lexicographical preferences virtually eliminate problems of indecisiveness, but they put a great deal of emphasis on the first-ranked variable, since it will determine choice unless a tie occurs between alternatives. Hence, when the first-ranked variable can take a wide range of values (particularly when it is an interval variable), this model may not generate significantly different predictions than one based on a unidimensional utility function.

Whether they are compensatory or noncompensatory, multidimensional models of utility need to predict in some general fashion the weights, thresholds, or orderings an individual will place on different utility variables in order to ensure the decisiveness of the resulting general model. It is clear that different individuals put emphasis on different goals, that goals often change over time, and that this has significant implications for their behavior. Moreover, these variations are not readily explainable as instrumental reactions to variations in the structural constraints surrounding them, which is one reason why conventional rational choice has generated so many anomalies.

Models of Time-Based Preference Formation

The growing literature on “time preferences” is aimed at correcting deficiencies in conventional rational choice assumptions about how individuals calculate overall utility when the effects of action are distributed over time. As was noted in chapter 1, conventional rational choice assumes that overall utility is a discounted sum of utilities for each future period, with the weight placed on the utility for a particular period being a geometrically declining function of the distance of that period from the current period. This has the convenient

---

172 The last is used in Tversky’s elimination-by-aspects model: see Tversky 1972. A modified version, where variables are linked in a tree structure, is presented in Tversky and Sattath 1979.

173 The original work in this field is Strotz 1955-56. For broad surveys and discussions of time preferences, see Loewenstein and Elster 1992 and Elster 1985.
property of ensuring that the relative weights on the utilities for any two future periods remain constant, even as time advances.

When this assumption of geometric discounting is relaxed, the relative weights placed on utility gained during different future periods may shift systematically over time, hence creating inconsistencies in the types of actions an individual wants to choose ahead of time and those he or she will actually choose.\(^{174}\) This in turn may provide individuals with an incentive to constrain their own future actions, even if this reduces the utility they gain in the current period. Hence, nongeometrical discounting of the future can account for the otherwise anomalous phenomenon of self-management or self-commitment, where individuals attempt to “bind” themselves in such a way as to limit the effects of time-linked preferences changes on their future actions.\(^{175}\)

Alternative formulations have been shown to do a better job than geometric discounting in predicting the way individuals in experimental settings choose to distribute a fixed level of benefits over time. Ainslie presents evidence that the rate of discounting declines as a function of time from the present, and he proposes a hyperbolic discounting function,\(^{176}\) while Loewenstein and Prelec present evidence that individuals, among other things, differentially discount gains and losses and seek to evenly distribute utilities across periods.\(^{177}\) As with other modifications to the conventional rational choice model, the decisiveness of any time preference model will be affected by the number of new variables it introduces to replace the geometric discount rate, as well as the precision with which the values of those variables can be a priori. While no consensus has been reached on the type of assumptions that ought to replace geometric discounting, a consensually held, parsimonious, and fairly precise general specification is perhaps closer at hand in this literature than in the other literatures surveyed in this chapter.

However, while they shed light on important phenomena, time preferences do not straightforwardly lend themselves to addressing the main indeterminacies and anomalies of applied conventional rational choice. While some types of behaviors that go against short-term wealth maximization may be accounted for by modifying conventional assumptions to assume that individuals have a highly negative discount rate for future wealth, this can be done within a geometric discounting

\(^{174}\) Of course, conventional geometric discounting assumptions also imply a kind of inconsistency over choices of actions, since individuals place no weight on utility gained during past periods and therefore regret any act of past consumption that reduced current or future utility. The main difference is that this has no bearing on action, since this inconsistency is retrospective, and the past actions cannot be altered. One could conceivably “correct” this inconsistency by specifying discounted utility functions that always start from some fixed anchor period, regardless of the current frame of reference. But this would generate strange anomalies of preference, since then a 100-year old man would have to be much happier about the enjoyment he had when he was an infant than about his current enjoyment.

\(^{175}\) Ainslie 1975; Thaler and Shefrin 1981; Schelling 1984; Elster 1979, chap. II.5.


Choosing an Identity

model. Furthermore, nothing in the empirical literature on time preferences would suggest highly negative discounting. Indeed, the most intuitive explanation for the majority of such behaviors, the existence of other-regarding preferences, has to do with an underlying utility function that applies to each period, not to the way in which utility is discounted across periods.

Models of Consumption-Based Preference Formation

The most prominent set of alternative assumptions about preference formation in the economics literature can be found in various models that posit that the utility an individual gains from consumption of particular goods will be a function of past consumption of those goods. Much of this literature focuses on addiction effects, as pioneered in a series of articles by economists, such as Pollack, Von Weiszacker, Stigler, Becker, and Winston, beginning in the early 1970s. Most rational choice addiction models assume a monotonically positive causal relationship between past consumption of a particular good and the present utility of consuming the good. Given this, the models can straightforwardly be adapted to the analysis of satiety if the sign of causality is reversed.

An alternative to assuming addiction or satiety when specifying a relationship between past consumption and current preferences is to posit that the utility gained from a good is partially a function of the magnitude of consumption of the good, but also of the rate of change in that magnitude over time. This is the approach taken in Scitovsky’s *Joyless Economy*, in which he labels the former “comfort” and the latter “pleasure,” suggesting that there are differences between individuals in the extent to which they value comfort as opposed to pleasure. His model draws from Berlyne’s work on hedonic arousal, which has been formalized more recently by Maital.

---

179 More explicit and extended analysis of satiation using rational choice models can be found as well. Margolis’ model of altruism is in effect a model of satiation, since it assumes that that the relative utility weights placed on either personal or group welfare are adjusted in the opposite direction to the relative provision of each in the past. See Margolis 1982, chap. 4. The social exchange models of Homans and Emerson discussed in the next section also incorporate a satiation effect as one component of their models of preference formation. Hirschman’s analysis of cycles of consumption makes predictions that are consistent with a satiation model, though it is based on the assumption that individuals systematically overestimate the utility of certain types of consumption and hence reduce the estimated utility attached to them once they are able to observe the real effects. See Hirschman 1982, chap. 1.
180 Scitovsky 1976, chap. 4. See also the retrospective on the book contained in a special issue of *Critical Review* 10, 4 (Fall, 1996).
Yet another alternative is Solomon and Corbit’s model of “opponent processes.”¹⁸³ In this model, which has been formalized by Taylor,¹⁸⁴ each act of consumption, in addition to contributing a short-lived primary utility of pleasure or pain, sets off a secondary process that generates utility of the opposite sign and decays more slowly over time. The intensity of the secondary process increases as a function of cumulative consumption. Opponent processes can be related to addiction because individuals may “redose” themselves with increasing amounts of a good in order to offset the secondary effects of earlier consumption. Clearly the direction of these effects is contingent on a number of factors, including the precise relationship between levels of consumption and utility, the way in which future utility is discounted, and the availability of alternative forms of pleasure.

Models of preference formation based on past consumption have the great asset of incorporating a dynamic relationship between past actions and current preferences, a necessary part of any intuitively plausible model of preference formation. However, for the models to be decisive in their predictions, they need to specify the shape and coefficients of the functions that transform the past consumption of relevant goods into a current utility weight. For the models to be general, they need to specify this for all goods that could conceivably enter an individual’s utility function and provide a clear set of criteria for categorizing different goods. For them to be accurate in their predictions, they also need to account for the clear differences that exist between goods in the nature of these functions as well as providing some specification of the original utility weights for each good for which preference change occurs. Furthermore, they must specify whether these effects differ between individuals, independent of past consumption, and, if so, how. Given this, generality, decisiveness and accuracy seem difficult to achieve for such models within a reasonably parsimonious set of assumptions.

A slightly unorthodox approach to addiction is taken by Becker and his colleagues,¹⁸⁵ who assume stable underlying utility functions and endogenously changing production functions that transform the consumption of a particular good into pleasure. However, other than

---


neutralizing the issue of time preferences,\textsuperscript{186} this does not make the
model more decisive than other consumption-based models. The most
obvious effect of such an approach is to turn questions about actors
(the nature of utility functions) into questions about structure (the
nature of production functions). Moreover, the model is implicitly
based on a maximand (pleasure) that is highly abstract and difficult to
operationalize and in fact not clearly differentiated from utility itself.
As was discussed earlier, this, when generalized, leaves the model highly
indecisive unless the pleasure production function for each and every
good is clearly specified.

One factor preventing the generalization of consumption-based
preference change models is that a focus on consumption tends to limit
the kinds of activities that can be modeled in a straightforward manner.
Particularly when processes of preference change are psychological
rather than physiological, the context in which a good is encountered
or experienced is likely to be equal or greater in importance to
preference change than the amount of the good consumed. Although
it has been argued that consumption models can be applied to
phenomena such as customs and traditions\textsuperscript{187} and romantic love,\textsuperscript{188}
making reasonably decisive and accurate predictions will require the
models to take into account the interactions between individuals,
goods, and the surrounding environment that determine actions and
outcomes. Specifying the nature of such effects would require a
fundamental transformation in these models. For instance, specifying
the “addictiveness” of romantic love would require taking into account
the personality characteristics of the individuals involved, the settings
in which their interactions occur, and the attitudes and actions of the
people around them. Such a transformation would clearly strain the
plausibility of using consumption, even metaphorically, as a description
of the action taking place.

\textit{Social Exchange Models of Value}
Like conventional rational choice, the social exchange approach in
sociology is based on the assumption that individuals choose actions
in order to obtain desired outcomes. However, it focuses primarily
upon repeated bilateral exchanges between individuals and typically
incorporates social goods such as status, approval, and affection, in

\textsuperscript{186} Because preferences over goods reflect a stable underlying utility function,
actors will always use preferences over goods for each future period rather than
present preferences when calculating expected utility. Whether such a model leads
to more accurate predictions than a utility-change model will presumably depend
on the nature of the good in question. Most of the empirical work on the Becker
et al. model has been on cigarette addiction, and the stable utility addiction
model is intuitively more plausible and easier to interpret clearly when applied
to physiological addiction than to psychological addiction. See Becker, Grossman,

\textsuperscript{187} Stigler and Becker 1977, 201-5.

\textsuperscript{188} Solomon, cited in Scitovsky 1976, 128-29.
addition to wealth, into individual preferences.\textsuperscript{189} The latter focus in particular has led social exchange scholars to examine in detail the value/utility that individuals place on different goods:

An early, incipient social exchange analysis of value can be found in the work of Homans.\textsuperscript{190} Homans’s overall model of action, though largely consistent with rational choice,\textsuperscript{191} is also heavily influenced by work in psychological behaviorism. Hence, “utility" is replaced with “reinforcement” and “action” with “emmission of activity.”\textsuperscript{192} Despite his behaviorist influences, Homans seeks to separate values from behavior. He recognizes the potential circularity of determining values from behavior, and hence the need for a priori determination.\textsuperscript{193} However, he does not attempt to provide a decisive model of preference formation, instead listing a set of factors that might be included in such a model.\textsuperscript{194} First of all, he implies, somewhat obliquely, that value can be inferred from the objective interests of the individual.\textsuperscript{195} More unequivocally, the value placed on a good is posited to increase as a function of the length of time an individual is deprived of it, a kind of satiation effect.\textsuperscript{196} Continuing, he argues that individuals’ values can be influenced by their “genetic past,” “ancient social history,” “social experience of members of a particular society,” and by each individual’s “unique experience.”\textsuperscript{197} Later still, he adds that value depends on the perceived justice of an interaction, which in turn depends on equality in the ratios between the costs and benefits of the various participants.\textsuperscript{198}

Blau’s subsequent analysis of exchange and power rejects the conventional rational choice view of values as self-interested and materialistic,\textsuperscript{199} asserting rather that “Most human pleasures have their

\textsuperscript{189} Surveys of recent work in the field can be found in Cook, O’Brien, and Kollock 1990 and Molm and Cook 1995. A broad collection of writings on the subject is Cook 1987. In recent years, the social exchange approach has placed less emphasis on rationality assumptions, in keeping with the increasing influence of social psychology within the approach. However, these newer models maintain the notion of a goal-oriented actor, and their assumptions remain consistent with some form of rational decision-making.

\textsuperscript{190} Homans 1958; 1961.

\textsuperscript{191} One difference is that individuals are seen as boundedly rational, avoiding long-term calculations and considering only readily available alternatives. See Homans 1961, 79-82.

\textsuperscript{192} For a wry comment on this terminological juxtaposition, see Heath 1976, 4.

\textsuperscript{193} Homans 1961, 41.

\textsuperscript{194} Homans 1961, 47-49.

\textsuperscript{195} For instance, he states that “...we have reason to believe, because we know that he is unskilled at his job, that help is a reinforcer, a reward, to Person.” See Homans 1961, 43. On pp. 44-5, he qualifies this by saying there may be occasional exceptions.

\textsuperscript{196} Homans 1961, 42-44, 54-56.

\textsuperscript{197} Homans 1961, 45-46.

\textsuperscript{198} Homans 1961, 72-78.

\textsuperscript{199} Blau 1964, 18-19, 236.
Choosing an Identity

roots in social life.” While he puts less explicit emphasis on values than Homans does, he has posited that all “social rewards” may be placed within six categories. These are divided along two dimensions. The first dimension distinguishes spontaneous evaluations, i.e., those that cannot be exchanged deliberately, from calculated actions. The other distinguishes between intrinsic rewards (those provided by the existence of a relationship), extrinsic rewards (those provided in exchange), and unilateral rewards (those provided in only one direction). The spontaneous rewards are personal attraction (intrinsic), social approval (extrinsic), and respect (unilateral), while the calculated ones are social acceptance (intrinsic), instrumental services (extrinsic), and compliance (unilateral). Respect, when transferred, becomes prestige for the recipient, while compliance becomes power. However, Blau does not discuss how these categories might be operationalized, nor does he specify the relative priority or weight that an individual would place on each type of reward. Later on, he discusses a very different sort of value, legitimacy, which is the shared value that groups of individuals place on the larger social institutions that determine the context of exchange. Nor does he rule out the existence of other values that exist outside of social exchange, though he argues that these values are less important, even “pathetic” in comparison.

Unlike Homans and Blau, Emerson not only analyzes the origins of values but explicitly attempts to account for how they change over time. In his theory, individual utilities are divided into a predesignated set of valued domains. The relative value that individuals place on a particular domain rises as the ratio of its level of consumption with levels in other domains falls. This implies a satiation effect for consumption within each particular domain. This in turn is posited to lead to an equilibrium in which levels in each domain are at some optimal ratio with one another. The optimal ratios are not specified, but values are closely linked to another concept that Emerson refers to as “need.” He posits that a domain “is needed to the extent that continued performance of activity and acquisition of other valued outcomes are contingent on acquisition of [it].” This, however, specifies the need for one domain in terms of the need of others. Without a clear anchor in domains whose need can be specified (at least partly) independent of the need for other domains, or a way of ascertaining optimal ratios directly, the model can be seen

200 Blau 1964, 14.
201 Blau 1964, 99-100.
202 Blau 1964, 270-71. The obverse of legitimacy is oppositional ideology. Two other types of shared values are the particularistic and universal, which can be seen as aggregated versions of some of the individual-level rewards described earlier.
204 Emerson 1987, 28.
205 Emerson 1987, 33.
206 Emerson 1987, 41.
207 Friedman 1987, 53-54.
as somewhat circular. If only a single highest-priority domain such as wealth is specified as having such an independent need, then this model reduces to one with a unidimensional value assumption, where nonindependent domains are seen as instrumentally connected to the independent one rather than being needed intrinsically. While Emerson does not rule out nonmaterial factors determining ultimate needs, he does not specify what such factors are.\footnote{208}

This social exchange literature, while providing complex general frameworks for the analysis of exchange value, has yet to provide a decisive model of value (preference) formation.\footnote{209} Homans analyzes several factors that influence value but does not build any systematic generalizations out of this analysis. Blau provides a categorization of different social values but does not account for the weight put on different categories. Emerson’s model more clearly specifies the value placed on different goods but does so in terms of their relationships with other values, while the origins of these other values is not specified. Satiation effects are also posited in both Homans’s and Emerson’s analyses, but the precise nature of these effects for particular goods and individuals is not specified.

No doubt future analysis of value in exchange theory will build greater determinism on the ideas of earlier authors. However, one would also expect that such analysis, while relaxing considerably the preference assumptions of conventional rational choice, will continue the tradition of focusing on value as a kind of exhaustible commodity gained in social exchanges, hence eschewing consideration of purely altruistic values (the willingness to give to another even in the absence of the other’s knowledge) and solitary expressive values (the desire to perform certain actions for their own sake, regardless of the social consequences).\footnote{210}

Conclusion

This survey of alternative methods for reforming conventional rational choice has shown that there is a wide range of alternatives for adding to or modifying conventional rational choice assumptions. With the exception of most of the structural models, all of the models in their most general form can conceivably be applied to analyzing human phenomena under any set of environmental conditions. Most of the models are also fairly parsimonious in the sense that their main assumptions can be listed and analyzed in the short space available in this chapter.

\footnote{208} It is possible that Emerson meant to do so at some later time. The paper remained incomplete at the time of his death in 1982.


\footnote{210} Blau seemingly rules out consideration of particularly the latter type of value, arguing that almost all significant value is gained from social life. He compares solitary eating to the “sparkling” social dinner, arguing that the former is “somewhat pathetic” by comparison. See Blau 1964, 14-15.
On the other hand, many of these models also introduce new variables for weights, thresholds, and orderings without providing a general way to predict their values. On the other hand, allowing ad hoc assumptions about these variables in midrange theories, threatens the overall decisiveness of the models. This makes them difficult to use to generate hypotheses across a wide range of environments. This is the case with all the nonoptimizing decision-making models discussed and all the models of multidimensional utility as well as social exchange models of value and most time preference and consumption-based preference models. Another common reason for indecisiveness in the models is the abstractness of the goods they place in utility functions, which makes their levels difficult to measure. This is the case with models of rational preference, some of the consumption-based preference models, and any model of inclusive fitness that seeks to take such preferences beyond family altruism.

The major advantage of all the models covered is their ability to accurately predict some set of behaviors that are anomalous for the conventional rational choice approach. However, some of this is due to their indecisiveness. When key variables in the models are left unspecified, the resulting theories can accommodate anything from action nearly identical to that predicted by conventional rational theories to the most extreme form of divergence, depending on the ad hoc assumptions adopted. The very existence of anomalies for conventional rational choice, despite numerous attempts to patch them up, is evidence that the general model is decisive.

Of course, some degree of loss in generality or decisiveness may be justified in order to gain greater accuracy. Moreover, there are also numerous settings in which even the conventional rational choice model is indecisive, particularly those in which it is implausible to assume that individuals have anything close to enough information to calculate expected utility or believe that one action dominates all others in generating wealth.

Finally, the increased accuracy of most of the unconventional models arises not simply from the looseness of their assumptions but also from the fact that they have systematically incorporated into their assumptions insights from empirical literatures on preferences, beliefs and actions, a reform that is essential if the shortcomings of the conventional model are to be repaired.

The next chapter presents my own attempt to address the shortcomings of the conventional rational choice model. It focuses on a new model of preference and belief formation, in particular on the underlying and relatively stable preferences and beliefs that precede various framing and discounting effects. Such a focus was not chosen because other types of reforms in the conventional assumptions are unimportant but because I believe that predicting preferences and beliefs is the key to addressing the major indeterminacies and anomalies of the conventional model in predicting economic, political, and social behavior.

None of the structural or decision-making models discussed in this chapter has been shown to provide much leverage in making
predictions on issues such as long-term government planning choices when information on future circumstances is scanty or nonexistent, or the circumstances under which individuals will participate in collective action when the costs to themselves far exceed the benefits. The intuitive explanation for such behavior is the existence of ideological or culture-based beliefs or of altruistic or discriminatory preferences. Furthermore, the preference and belief models surveyed either do not address these types of issues, or do so in a way that is incompletely specified at the general level and therefore difficult to use in making decisive predictions.

This does not imply that these models are not useful or that some version of them should not eventually be adopted as part of a general reform of the conventional rational choice model. Indeed, while they may be less applicable to the kinds of empirical issues discussed here and later in the book, they may often be of equal or greater use when applied to empirical issues in other arenas or at different levels of analysis. Furthermore, any thorough reform of the conventional rational choice approach ought to take place in multiple areas. While the focus in the next few chapters is on preferences and beliefs, I discuss broader reform of the rational choice model in chapter 7, drawing from many of the models discussed here.

As was mentioned earlier, I have limited my survey of preference and belief models to those whose assumptions have been stated in a general and choice-theoretic form and can therefore be straightforwardly incorporated into the rational choice model. However, what is perhaps most striking about the models covered is the fact that they bypass the largest single category of preference and belief models in the social sciences. These models assume that actors are engaged in a collective process of constructing their own identities and that this process is aimed at creating an individual and collective sense of self that is both positive and coherent. Rather than considering preferences, beliefs and actions in isolation, they assume a mutually causative relationship between these entities.

One reason why the commonality between these models is seldom recognized is that they exist under a wide variety of labels, including social identity theory, the social construction of reality, dissonance theory, and coherence theory, or are sometimes simply implicit in the basic assumptions of an entire field of study, such as postmodern theory or symbolic interaction, new social movements, neo-nationalism, and the new institutionalism. Furthermore, there is often great debate over the definition of concepts, as well as the relative emphasis to give to individual versus collective identity or to positive versus coherent self-images. They are drawn from a wide range of disciplines, primarily psychology, sociology, and philosophy but also anthropology, political science and even economics, and tend to be couched in very different terminology and evaluated under very different sets of methodological criteria. Nonetheless, I attempt to show how the basic assumptions of all these models are essentially compatible with one another and with individual rationality, even those drawn from traditions quite distant from the choice-theoretic tradition discussed in this chapter.
I also attempt to show how a fairly simple formal model can address some of the theoretical disagreements that exist within these literatures, showing how seemingly conflicting sets of predictions about preferences and beliefs are due to differences in assumptions about environments rather than differences in assumptions about the underlying processes through which preferences and beliefs are formed.
Chapter 3
A General Model of Preference and Belief Formation

In this chapter, I present a general model of preference and belief formation. It is complementary to the rational optimization model of decision making, and is integrated with it to form the basis for a revised general model of action. In the model, choice of action is seen as the result of a decision-making process aimed at maximizing preferences given beliefs, but also triggers changes in these preferences and beliefs.

The next section of the chapter defines the concepts of decidability, doubt, and coherence that are central to the model, then discuss the model's basic assumptions. The remaining sections analyze preference and belief formation, examining static, dynamic, and collective choice contexts. These sections include implications of the model for particular empirical phenomena and theories about them.

The basic ideas contained in the model are hardly new—they are instead an integration of existing ideas that have been drawn from a variety of academic fields. The ideas are associated with work on individual attitudes, identity, and agency but tie them to a work on collective ideology, culture, and norms. The theoretical literatures involved span all the major social science disciplines and philosophy, and the discussion often juxtaposes the work of academic communities that rarely engage in dialogue. I hope to show that the model does more than stack an eclectic bunch of concepts and theories on top of one another, that it also provides a way to tie their common ideas together, to isolate and address their areas of disagreement, and to extend their implications.

I do not claim that this model reflects anything close to the entire reality of human motivation and cognition. Nonetheless, I try to show that it can provide a general and parsimonious way to account for preferences and beliefs and that its predictions are useful determinate and conform with broad patterns of empirical evidence. I also try to show that it ties empirical patterns to deeper underlying processes and generates novel hypotheses for future investigation. In chapters 4 through 6, I try to show that the model of action that results from integrating the model with a rational optimization model of decision making provides a useful tool for analysis and theory-building on substantive issues that existing rational choice and nonrational choice theories have failed to address adequately.

I should note that the model does not specify the bounds that limited or biased information processing places on an individual's ability to optimize. This does not reflect any opposition on my part to bounded rationality models. Instead, I argue that boundedness is consistent with and in some ways even necessary for optimization. In the absence of knowledge that is both infallible and complete, optima can be defined only with respect to constraints that are subjectively constructed around past experiences. Not only will consideration of alternatives be inherently limited, an expansion
in one’s choice framework that is unnecessary to accommodate new experiences may hinder rather than facilitate optimization. Even so, I do not pretend to provide a fully adequate model of boundedness, since the main emphasis in this chapter is presenting new assumptions about preference and belief formation, and specifying the location of other important decision-making bounds on top of this would be an unrealistic task. In chapter 7, there is a brief discussion of how the preference and belief model presented here can be integrated with satisficing models and models of nonexpected utility maximization.

An Identity Coherence Model of Preference and Belief Formation

Formal specification of the model is based on a subjective probability framework, though one quite modified from that originated by Savage and widely developed elsewhere. The notation is hybrid, imparting a flavor of formal logic to traditional choice-theoretic symbols. Some reasons for selecting the notation will be provided as the model is presented, and clarifications are provided when interpretation might be ambiguous.

An individual’s preferences and beliefs are represented by a rule base $P$, consisting of a set of rules $p_i = (s_i, b_i)$, where $s_i$ is a deterministic statement regarding the state of reality and $b_i$ is a probability weight in $[0, 1]$. A constant within a rule base is the name of a discrete object, an ordered or unordered class of objects, a cartesian product of classes, or a mapping between classes. A variable is an expression whose denotation is within the class but is not fixed.

Variables come in multiple flavors, depending on the scope of their denotation. A contingency is a definite description, i.e., a variable with a unique denotation that is state-dependent. Contingencies are variables in the algebraic and choice-theoretic sense, and will implicitly be quantified as unique existences. As per usual choice-theoretic convention, a contingency is represented by a lowercase symbol. To maintain consistency, this usage will extend to quoted and unquoted propositions as well. The entity denoted by a contingency in a particular state will be called its disposition in that state. Wild cards represent every entity within a class, and are implicitly quantified universally. A wild card is represented by a lowercase symbol with a dot or umlaut over it, e.g., $\dot{x}$, $\ddot{x}$. Contingencies over the same class of entities may themselves come in classes, and these are distinguished by subscripts. In this case, the text symbol will be an attribute.

---

1 See Savage 1951; 1953. Prominent surveys include Kreps 1988, chap. 8 and 9; Fishburn 1988, chap. 1; Fishburn 1970, chap. 10.
2 In Kripkean parlance, constants are rigid designators, while variables are Fishburn sic designators. See Kripke 1980, chap. 1.
3 Russell 1918.
4 Contingencies are sometimes referred to as (nonary) functions in computational applications of predicate calculus, and are called propositional variables when they range over truth values.
A General Model of Preference and Belief Formation

i.e., a nondefinite description, and subscript(s) will denote one or more entities, such as individuals or points in time, to which the attribute is indexed. Variables appearing in subscripts are wild cards.

In part because of possible confusion over different senses of the term variable as well as that between values as preferences or as assignments to variables, these terms are avoided for the most part in this chapter in favor of the terms contingency, wild card, and disposition. Words may be used instead of letters of the alphabet when the model is applied to a real-world scenario.

A rule base imposes a set of subjective constraints on possible realities, and generates a set of candidate solutions for the disposition of any contingency. For a contingency \( x \), I use the circumflex notation \( \hat{x}(P) \) to represent the set of dispositions for \( x \) that are implied possible by \( P \), i.e., \( \{ \hat{x} : Pr(x = \hat{x} | P) > 0 \} \). I use the notation \( \hat{x}(P) \) to represent a contingency over \( \hat{X}(P) \), while \( \bar{x}(P) \) represents the expected disposition of \( x \) given \( P \). Any contingency in a rule base whose disposition is not specified determinately is called a fate.

Assumptions of the Model

I begin with a few assumptions about rule bases, then move forward to assumptions about preferences and beliefs.

Assumption 1: Each individual’s rule base is finite and consistent, and statements are expressed in causal form.

The first two parts are straightforward. Finiteness refers to the number of statements present in \( P \). Consistency means that \( P \vdash \bot \) is invalid.

Causal form restricts statements to those of the form \( x = fy \), where \( x \) is a contingency and \( fy \) is an expression over a (possibly empty) set of other contingencies \( y \). Each contingency in \( y \) is seen as a cause of \( x \), and causality is seen as a transitive and asymmetrical relationship, though not necessarily complete. There is no assumption (indeed, things are necessarily otherwise) that all causal rules express the metaphysical relationships that social scientists would define as causality, only that their form is causal. This allows us to partition a rule base into subsets \( P_x \), where each \( p_i = (s_i, b_i) \in P_x \) has \( x \) as the contingency whose causal model is being specified. There is no assumption that \( p_i \) are disjoint, specify a complete probability distribution, or even have the same parameteric form, i.e., they do not have to all claim to represent direct causality, as long as their causal orderings are consistent.

I call the ordered set of contingencies attached to the left-hand side of statements within \( P \) the vocabulary \( S(P) \) of \( P \). The set of possible dispositions over his or her vocabulary is \( \hat{S}(P) \), and a contingency over \( \hat{S}(P) \) is \( \hat{S}(P) \). The language \( P \) is the set of statements that are deemed possible by \( P \), e.g., \( \{ s : Pr(s | P) > 0 \} \).

Assumption 2: Each rule base is self-defining and self-referential. Self-definition means that all criteria that determine the form and content of an individual’s rule base is represented within the rule base itself. The assumption of self-definition is useful because it allows us to avoid taking an individual’s language for conceptualizing reality as
fixed or exogenous to the preference and belief formation process, but rather as part of that process.

**Self-referentiality** means that $P$ itself is in the vocabulary of $P$ for all individuals. An individual may view rules in $P$ as being generated by contingencies, and as generating others. The assumption of self-referentiality is useful because it allows us to represent an individual’s choice sets as ordinary statements in his or her rule base. It also implies the ability of an individual to empathetically use the vocabulary of his or her own rule base as the basis for a model of her own past and future rule bases as well as the rule bases of others.

Self-definition and self-referentiality together imply self-consciousness, which requires that all of an individual’s decision-making and deliberational algorithms exist in the form of beliefs about his or her own past and future actions, preferences, and beliefs.

One issue that should be addressed is whether a self-defining rule base is possible, and, if so, how such a rule base could be created and what its content will be. This requires a set of basic assumptions about inherent human representations and learning. I will attempt keep these as minimal as possible, and to derive most of my suppositions as theorems.

The primitive concepts in rule bases are a set of discrete sense contingencies $\Sigma$, with each $\sigma \in \Sigma$ an atomic proposition. This is the union of a set of external contingencies $\Xi$ generated by environmental stimuli and a disjoint set of internal contingencies $\Upsilon$ generated by his or her own thoughts. Any string of sense contingencies $\sigma$ is also called a sense contingency. An individual’s experiences are initially viewed as a stream of sense states $\sigma_t \in \Sigma^* = \bigoplus_{i=1}^{\infty} \Sigma_i$ comprising strings of arbitrary length drawn from the members of $\Sigma$. The sole primitive operator is salience, $\phi$, which maps each pair of sense contingencies onto $[0, 1]$, and represents the proportion of the occasions in which the first sense contingency has been detected that have immediately been preceded by detection of the second.

While assuming no innate ability to communicate verbally or assign symbolic meaning to sense contingencies, nor to recall specific past sense states, I make the informal assumption that people are innately able to distinguish between external and internal sense contingencies as well as among external contingencies corresponding to shapes, sounds, and smells. Furthermore, while they will be able to generate and manipulate internal sense contingencies voluntarily, other internal contingencies will also be generated reactively through interaction with external contingencies, including a secondary image of sense contingencies salient to those present in the current state and affective responses to both the current sense state and salient images.

I label as a sense-defined concept any sequence of external sense contingencies $\xi \equiv a$ that forms a potential symbol, e.g., a vocal sound or simple shape, and that has (nearly) always been preceded by a sense contingency $\sigma$ such that $\phi(a, \sigma) \approx 1$. The initial learning of words and meanings thus occurs through a process of salience-building through the juxtaposition of a sense contingency with another that becomes its symbolic representation.
Individuals are also able to make other kinds of distinctions based upon whether the sense contingencies in question are internal or external. An internal sense contingency $\nu$ that follows from an external one $\xi$, such that $\phi(\xi, \nu) \approx 1$, is an affect, while an internal sense contingency that is succeeded by an external one such that $\phi(\nu, \xi) \approx 1$, is a volition. In the latter case, the corresponding external property $\xi$ is an action.

It is clear that a rudimentary set of associations between concepts and sense contingencies can be built up in this fashion. However, self-consciousness implies that the salience will itself be be associated with the linguistic operator “=” i.e., “is,” since it will initially nearly always be heard with regards to an internal sense contingency of high salience between two concepts. Once such an association is set up, more abstract levels of generalization may occur, since now the communicated statement $x = a$ is known to be equivalent to the internal sense contingency $\phi(x, a) = 1$. I will call propositions that are not sense-defined but rather are defined through linguistic association as idea-defined concepts, and they are one of a larger set of ideas that can be generated through communication.

Moreover, since internal sense contingencies correspond to an individual’s own thoughts, beliefs about such conditions constitute beliefs about his or her own rule base. If an individual knows that the salience corresponding to the string of external sense contingencies $x = a$ is $\phi(x, a)$, he or she can represent this salience as the existence of the statement $x = a$ in $P$. Hence self-definition can be seen to enable self-referentiality.

If a statement is a subjective volition for an individual, that person will believe he or she has total control over its truthfulness. Due to self-referentiality, this is equivalent to believing that one’s own belief in a statement will cause it to come true, i.e., statement $s$ is a volition for her then $(s, 1) \in P_{\text{top}}(s) = 1$. A statement that is both held $((s, 1) \in P)$ by an individual and seen as a volition at the time he or she holds it will be an intention, whether or not she continues to see it as a volition later. The proposition $p(s) = 1$ expressed by an intention $(s, 1)$ will its strategy, and also an action if the intention is currently seen as a volition. A contingency $x$ is volitional for an individual if for all $\hat{x}$, $x = \hat{x}$ is a seen as a volition. A proposition $p$ is an outcome of $(s, 1) \in P$ if and only if $s$ is an intention and is both necessary and sufficient to derive $p$ given other rules in $P$, i.e., $(s, 1) \in S$, $P \setminus (s, 1) \vdash p$ is invalid, but $S \vdash p$ is valid.

In general, an individual defines volitions and outcomes (including actions) only in terms of her present rule base, while intentions and strategies are defined relative to a rule base of an individual and/or point in time. For example, an individual may think he or she was mistaken in previously viewing a particular statement as a volition and about the outcomes arising from adopting it as an intention yet continue to define it as a past intention. Another way to put this is to say that statements regarding intentions and strategies are metasubjective, i.e., subjective views of some other’s (including past or future self’s) rule base, while those about volitions and outcomes are not.
An individual’s set of recognized sense contingencies is a private vocabulary in the sense that he or she is not able to communicate any of these ideas directly to others. However, an individual can use her public language of concepts to communicate certain sense contingencies, with each constant superimposed on a sense contingency. Moreover, the learning of a public language of symbolic representation allows an individual to represent specific sense states. Indeed, the concepts an individual learns through communication determine how he or she organizes sense contingencies into categories and remembers past experiences.

I use the notion of private vocabulary rather than private language to sidestep the question of whether language is possible independent of social interaction. The existence of a vocabulary simply indicates that an individual has a way of partitioning states in his or her representation, while a language requires a set of grammatical rules distinguishing well-formed expressions from ill-formed ones. Sense contingencies are discernable in that an individual can tell whether they have been present, but atomic in the sense that there will be no concepts within an individual’s private vocabulary with which to analyze their internal structure. The $\phi$ operator can be taken as existing prior to language, but it is simply a marker for an association rather than a syntactic tool that an individual can manipulate to generate new representations.

To these starters, I add some assumptions about how such concepts are translated into a rule base, how a rule base shapes action, and how preferences and beliefs within a rule base are formed. To begin with:

**Assumption 3**: Each individual needs to think that he or she is optimizing some purpose through his or her existence.

Utility, of course, is the purpose that an individual decides he or she is optimizing. Standing the typical definition of utility on its head allows us to view the problem of preference and belief formation as one of finding an appropriate rule base that allows each action performed to be seen as optimizing.

Absent constraints, however, the entire notion of optimization is ill-defined. If an individual sets a rule base so that an arbitrarily high amount of utility is possible under any possible state of the environment, then no actions will maximize utility. Hence, an individual who views himself or herself as omnipotent will somewhat paradoxically be unable to think that she is optimizing. However, an individual who views himself or herself as completely helpless will also have the same problem, since he or she will have no actions to perform. Hence rational optimization in this model will not consist of the avoidance of constraints, but the subjective construction of constraints appropriate to a teleological view of life.

In order to believe that he or she is optimizing something, an individual must have confidence that choices over actions are the best.

---

5 The controversy over private language arises from the Wittgenstein’s argument that a language must have a correct and incorrect usage, which in turn cannot be distinguished without social reference. See Wittgenstein 1953, sec. 243-315.
for whatever he or she defines as utility. This requires two things from a rule base: First, an individual must be able to calculate the utilities that be expected under every strategy for every disposition of fates that is implied possible by the rule base. Second, given this, there must be some overall strategy that maximizes the individual’s utility compared to other feasible strategies across each disposition of fates.

Analysis of preference and belief change and action therefore centers around three central concepts, *decidability*, *doubt*, and *coherence*. Decidability refers to the existence of an algorithm that allows an individual to determine optimal choices. Doubt refers to the expected gap between the utility provided by a chosen strategy and that provided by the optimal available strategy under each disposition of fates. Decidability and doubt may be analyzed in relation to choices over specific contingencies, while coherence is always be a property of a rule base as a whole.

Utility is represented by $u$. When different strategies are being considered for particular volitional contingencies, rules are held in suspension. For such analysis, expected utility generated under rule base $P$ under strategy $v = a$ will be represented as $\hat{u}(P_{-v}, v = a)$, where $P_{-v}$ is $P$ with the rules for $v$ absent and $v = a$ is equivalent to the rule $(v = a, 1)$, and where a comma represents the grafting of a rule onto a set, e.g., $P,p \leftrightarrow P \cup \{p\}$. I represent her utility under each disposition $S \in \hat{S}(P_{-v})$ of fates as $\hat{u}(S, v = a)$.

Doubt under such a strategy is represented as

$$\hat{u}(P_{-v}, v = a) = \sum_{S \in \hat{S}} (P_{-v})\hat{u}(S, v = a) - \hat{u}(S, v = v^*(S))p(S),$$

where $v^*$ is an optimal strategy among those available in an individual’s choice set of strategies $V$ for each disposition of fates $S$. Doubt is always nonnegative and is positive whenever there is any disposition of $S$ with positive subjective probability for which the utility provided by the optimal strategy is less than the maximum feasible utility in that state.

Clearly, decidability implies that doubt is defined, and the strategy that minimizes doubt will also minimize expected utility. Doubt exists whenever there is a perceived possibility that a person’s chosen strategy is not the best feasible one for meeting his or her goals. Either undecidability or doubt is posited to trigger a search for alternative strategies, preferences and beliefs.

Doubt is closely related to the concept of regret in economic and decision theories. Indeed, it corresponds formally with the expected level of regret if Savage’s specification is used.\(^6\) Alternatively, regret can be seen as the special case of doubt for a past action when the consequences of the action are known. However, the role of doubt in the model here and its substantive interpretation are quite different from that of regret in economics, where it is used a maximand for

---

\(^6\) Savage 1951.
decision-making models. Indeed, its use corresponds more closely with the use of concepts like dissonance, consistency, and congruence in social psychology. These all refer to some perceived problem in the fit within an individual’s attitudes and/or actions, one that generates a change in attitudes.

Decidability and doubt can be linked to strategies over one action or a series of actions. As in conventional rational choice models, analysis of some of an individual’s actions in isolation from earlier or later ones implies that actions in the choice set being considered will have no effect on the expected utility provided by actions in choice sets not being considered. However, it also implies that preference and belief change regarding the actions being analyzed should have no effect on doubt regarding other actions. In other words, analysis of local optimization of doubt implies its compatibility with global optimization in terms of longer-term specification of strategies and outcomes.

One important aspect of the calculation of doubt is that it is bidirectional in relationship to time. It can be both retrospective and prospective, as long as both past and future actions can affect an individual’s expected utility. Prospective doubt may be generated by a perceived stochastic element in the relative utilities generated by different actions, while retrospective doubt may also be generated by changes in the beliefs that led to previous actions. Overall, an individual’s cumulative doubt will be linked to strategies over all past and intended future actions, to an individual’s “life strategy.”

Since doubt is bidirectional, individual adjustment of preferences and beliefs take into account both past and future actions. Thus, preference and belief change will depend upon actions across both past and future periods. While conventional rational choice formalizations do not prevent utility calculations for past actions, these are ignored because they are irrelevant for current actions, and past actions themselves are relevant only to the extent to which they affect current structure. However, in this model, utility calculations for past actions (and actions foregone) can influence change in preferences and beliefs, and hence they will also influence the relative utility attached to future actions.

Of course, there is a key difference between past and future actions, since intentions can always be changed. This means that prospective doubt minimization can conceivably involve a change in preferences, beliefs, and strategies, while retrospective minimization can only involve change in preferences and beliefs. This in turn implies that past actions commit individuals to particular types of preferences and beliefs in a way that future actions do not. It also implies that, while future actions are always perceived to be optimal even when they generate doubt, past actions may be seen as suboptimal in the light of new information or attitude change.

Coherence describes a condition in which, given preferences and belief, an individual’s choice set over all periods is decidable and doubt over a chosen strategy is equal to zero. Within the subjective

---

7 To borrow a phrase from my mutual fund.
probability framework, coherence implies the existence of a strategy whose utility is equal to or higher than that of all other choices in all states with nonzero subjective probability. Substantively, it implies the perceived existence of a course of action, consistent with what an individual has already done, that is assuredly the best one for meeting her goals. The model as a whole is referred to as a coherence model.

While numerous definitions of identity exist in various psychological literatures, there is a shared notion of identity as the “unified, purposive, aspect of self,” something that is quite consistent with the definition offered here. Being individualistic puts the definition superficially at odds with conceptual trends in the sociology of identity, where identity is usually viewed as a socially constructed phenomenon. However, the definition and model built around it are quite compatible with social construction, the main difference being the explicit emphasis the model places on social context and identity.

The definition of coherence is also compatible with that in philosophy, where coherence usually refers to a relationship between an individual’s beliefs that renders these beliefs justifiable, but it is also used to refer to a more general relationship of this sort between beliefs, desires, and intentions. The notion of coherence is also consistent with the concept of narrative unity, which is used in philosophy to describe a life that can be described as fulfilling a clear purpose.

In order to analyze how rules are chosen, we can differentiate between those rules that are placed in an individual’s rule base exogenously through experiences and through initial learning of concepts, and those added later through deliberation and the endogenous generation and modification of rules. I call the initial rules an individual’s foundation and those that are added later as her identity.

Assumption 4: Sensory experiences as interpreted through learned concepts will comprise an individual’s foundation, and cannot be removed from a rule base.

While raw sense states cannot be held in memory except through their “residue” as saliences, acquisition of symbolic representation capability will allow an individual to recall specific states via statements in his or her rule base. Each statement is attached to a probability of 1 equivalent to the evaluation $T$.

An instance of a sense contingency that cannot be interpreted in this fashion will be forgotten. Moreover, a condition that is chunked together with others within a sense-defined concept may be remembered even if it did not occur, as long as its accompanying conditions did occur and there is no other way to interpret them absent a missing piece. Otherwise, an individual will not be able to forget or change the past experience, although he or she may be able to reinterpret at a higher level by altering the implication of the sense-defined concept for other concepts.

---

8 Frable 1997, 139.
This assumption can be labeled an information constraint assumption, in contrast to the information assumption of conventional rational choice theory discussed in chapter 1. The information assumption implies that beliefs, other than the common knowledge of rationality, consist solely of propositions that follow logically from observable characteristics of the environment. The information constraint assumption does not limit beliefs in this way but implies that all observations (i.e., sensory experienced interpreted in terms of sense-defined concepts) will be present in her foundation. Along with logical consistency, this implies the following:

**Proposition:** An individual’s beliefs will never contradict her observations, though observations will not fully determine her beliefs.

An assumption of this type is thus useful in establishing some correspondence between beliefs and experiences, and preventing individuals from eliminating doubt by constructing fantasy worlds. Furthermore, there are good reasons to presuppose that individuals will not have any reason to construct such a world. First, a foundation will never constrain an individual from achieving coherence since there will always exist some conceivable set of preferences and beliefs that can provide a purpose to a set of experiences. Rather than providing too many constraints, they leave optimal choices underdetermined. Furthermore, even if an individual seeks to change his or her foundation, there will be little basis for doing so, since there is no communicative means by which to gain new definitions for sense-defined concepts, nor a clear way to produce an alternative conceptualization de novo.

The assumption implies that all individuals hold two types of beliefs. Foundational beliefs describe observations individuals have made about the environment. Identity beliefs concern aspects of the environment that are not observable, including causal forces as well as past and present characteristics of abstract or hidden objects. Indeed, we can divide the set of rules for any contingency $P_x$ into foundational rules $P_{0,x}$ and identity rules $P_{1,x}$, where the former encompass experiences and sense definitions for the contingency, and the latter encompass causal rules, with the former acting as a set of metaconstraints on the types of statements that can be incorporated into the latter.

Part of the experiences that will be retained in an individual’s foundation will be memories of past actions. Furthermore, a foundations will also define sets of contingencies that are viewed as action, since each outcome that is always tied to an internal sense contingency will be defined as being under direct volitional control until such time as the feedback that this is not the case occurs.

While foundational beliefs may underdetermine rule bases, they also provide the raw materials from which an identity is constructed. The reason for this is that only a small percentage of the propositions contained in an individual’s foundation interpret directly to rules in his or her rule base. The rest remain latent because they are *embedded* propositions, i.e., recalled experiences that have symbolic propositional meaning but are not equivalent to the propositions they are expressing.
The largest set of embedded statements will be communicated ideas that an individual receives from other individuals but that are not initially juxtaposed with a sense contingency. Unlike sense-defined concepts, such ideas may be accepted or rejected, but their propositional content will be understood if the variables and constants contained within them correspond to those that are already a part of the vocabulary of the rule base. Furthermore, they will be retained in memory as quoted rules within statements. For instance, a statement $s$ received at time $t$ will be retained as a rule of the form $(m_t = 's', 1)$, where $m_t$ stands for “message received at time $t$.” Note that the statement portion of the rule is not $s$, but $m_t = 's'$, hence $s$ is not converted to a rule but embedded within one and left unevaluated.

Since they are latent but available, an individual can use communicated ideas as the basis for generating new rules. In doing so, however, he or she will engage in a deliberative process that may transform these ideas while generating rule candidates. These transformed ideas, along with the originals, will be retained in his or her memories as embedded statements, even if they are rejected as new rules, and will be resources for addressing later choice problems.

The set of embedded statements that are relevant to generating rules for a particular contingency are called the **choice frame** for that contingency, and include all statements that include the contingency within them. The set of embedded statements that are relevant to generating rules for any contingency within an individual’s vocabulary is called the **identity frame**. An individual’s choice set over rules for a particular contingency consists of the set of statements in the choice frame parsed using the rules of the current rule base and rendered into causal form. Likewise, an identity set consists of all embedded statements within an identity frame parsed in this manner.

A choice set over strategies includes those strategies that are suggested for her in embedded rules and also optimal over some identity that can be generated from embedded statements in an individual’s foundation. It is important to note that a choice set over strategies will even retrospectively contain only those strategies that would have been prospectively optimal for some available identity. In the most obvious case, an individual who has engaged in iterated random wagers does not have to compare his or her strategy retrospectively to that of guessing correctly every single time since neither such a strategy nor an iteration by iteration description of it was optimal under any available identity.

Embedded statements consist of idea-based definitions of one concept in terms of another, or of empirical or normative ideas. Normative ideas identify certain contingencies as desirable or undesirable through their links with utility, while empirical ideas contain propositions about the links between contingencies and conceptual ideas define certain contingencies in terms of others. The weight $p(s)$ for a statement $s$ can be thought of as an individual’s orientation towards the idea. Given this, all identity can be thought of as collected orientations toward conceptual, normative and empirical ideas.
Ideas arise from an individual’s cultural milieu, hence we can think of an identity frame as being structurally determined. However, because they can be rejected and manipulated, ideas provide only a framework that an individual can bend around his or her own unique experiences within her milieu. On the one hand, communication is simply a type of observation in which stimuli have decodable symbolic content. On the other hand, while sense-defined concepts provide no ready vocabulary for reconceptualization, ideas broadcast through communication often contain the seeds of their own negation.

Since the presence of an idea within an individual’s identity frame is determined by the communicative actions of others, communication is important for reasons beyond the information it may provide about the environment. Individuals may observe communications directed toward others by third parties, and thereby make inferences about the set of ideas available to both sender and recipient. Furthermore, individuals can infer from another individual’s cultural background what sorts of communications he or she might have received, even if they have not observed this directly. By combining knowledge of the ideas available to an individual with knowledge of his or her background, inferences can be made about attitudes. Furthermore, these inferences can have an effect not only on the actions of those interacting with and individual, but the types of orientations they themselves adopt toward ideas available to them.

Substantively, the assumption suggests a way to reconcile theoretical approaches that view individual attitudes as entirely a product of social learning and those that emphasize logical processes. It does not view individuals as passively accepting any idea because it is prevalent in their cultural milieu, nor as ignoring all ideas as mere propaganda or unreliable sources of information. Instead, individuals use ideas as toeholds for addressing their own internal optimization problems. However, if an idea does not help to solve these problems, individuals will reject it. Indeed, one of the implications is that communication of an idea is more effective in determining the dimension of attitude change than its direction, and can promote the formation of attitudes opposed to that expressed in the communication. A normative idea causes an individual to contemplate both valuing or disvaluing a particular type of good. Likewise, an empirical idea will cause an individual to contemplate both the existence and nonexistence of a particular state of the environment.

However, another chicken and egg issue arises. While experiences can define the initial concepts in an individual’s rule base, and even the salience of different concepts to one another, what accounts for an individual’s original behavior absent an initial utility function? Even if an individual has the ability to choose an action, on what basis will such a choice be made initially?

---

9 Despite the frequent tendency to contrast structure (environment) and culture, most of the variables viewed as structural in social science theories (classes, modes of production, etc.) have connotations that cannot be reduced to physical properties, and are thus cultural as well.
Assumption 5: An individual’s initial utility function will be equal to the salience-weighted sum of his or her hard-wired affects toward the sense contingencies linked by associations to his or her current sense state.

We need not assume that individuals are rational from birth. Indeed, absent the ability to represent an environment symbolically, it will not be possible to retain specific sense states in memory, and hence learning is restricted to various accumulated saliences between sense contingencies. However, prerational individuals will still engage in behavior based upon innate affects and operant learning, and their later selves will seek to rationalize such behavior. Each prospective behavior will have saliences linked to its past immediate consequences, which in turn will generate internal sense contingencies linked to the secondary images of these consequences in an individual’s mind. These sense contingencies will generate affects, which then drive behavior. However, such learning will be inherently short term in nature. Furthermore, saliences of consequences, based purely on normalized aggregation over past periods, will be inaccurate matches to the probabilities attached to them.

An individual who is just beginning to exercise symbolic competence will have to find a purpose for past behavior, which in turn will be largely affect-driven. Given this, the definition of utility that can best rationalize such behavior is one in which affect is seen as the source of utility. Indeed, absent deliberative intervention to generate new volitions, affects by definition must drive behavior. Subjectively, actions have been defined as external sense contingencies generated by internal sense contingencies, volitions. However, absent deliberation, no internal sense contingencies will be generated other than those, i.e., affects, that have been generated by external sense contingencies. It follows that absent deliberation the only volitions that can exist will be those that are also affects. And the only way that affects can be influenced by prospective future events, barring deliberation, are through the secondary images generated by saliences.

Given that initial affects are driven by physical urges and needs, this implies the following:

*Proposition:* An individual’s initial utility function will always reflect a positive weighting for her physical well-being.

As an individual gains in deliberative ability and is able to choose actions consciously, the inadequacy of purely affect-defined utility and salience-defined probabilities for attaining coherence will become apparent, and he or she will use her ability to consciously generate internal sense contingencies to actively begin changing her conceptions of utility and probability. Hence while utility and probability will initially be equivalent to affects and saliences, they will become increasingly deanchored from these origins as time goes on.

This assumption embodies, in highly simplified form, the notion that there is a hard-wired orientation toward short-term physical comfort and an aversion toward physically damaging outcomes. Because the goods necessary to influence such outcomes are generally tradable, it is possible to cover them with a blanket concept such
Choosing an Identity

as material welfare, which in turn can be measured through proxy indicators such as the monetary value of consumption goods. While it would be more realistic to incorporate maximands such as nurturance and status into hard-wired utility, this would introduce the problem of specifying a priori weights. Furthermore, such maximands are less tradable and specific to particular relationships that cannot easily be accounted for beforehand but can be accounted for as a product of preference formation.

The assumption provides an anchor to preferences, hence avoiding the prospect that individuals will start off with a yogic utility function, i.e., one that is indifferent over all outcomes. It is also a useful way of specifying the initial preferences that individuals bring to any choice situation being analyzed, providing a baseline from which further preference change can occur. It is clearly impractical in most applications to trace preference and belief change back to each individual's infancy. Moreover, the model will not work by assuming that individuals start off as tabula rasa, since there needs to be some set of preferences and beliefs to inform their initial expected utility calculations. Therefore, there needs to be some alternative specification of initial preferences and beliefs. Given analysis beginning at a particular period, we can impute initial beliefs by making structural assumptions implying that certain aspects of the environment are currently observable to individuals. However, such a technique cannot be used with preferences, and it makes sense to avoid making ad hoc assumptions about even initial preferences.

The alternative is to take the positive valuation of material welfare as a general assumption about initial preferences and account for other preferences endogenously by including the time period in which such preferences are thought to have come about within the analysis. This method closes the door to ad hoc preference assumptions, and sets the conventional rational choice preference assumption up as a baseline from which to analyze divergence, hence facilitating the comparison of predictions. The coefficient attached to material welfare is irrelevant if it is the sole initial variable since the implications of utility functions are invariant across positive affine transformations of transient utility.

An individual enjoying coherence will by definition be assured of the optimality of his or her actions across all possible futures. As long as no event with zero subjective probability occurs, coherence will be preserved. Hence, preference and belief change will be triggered by surprises. Given surprise, an individual will have to deal with two simultaneous problems. The first is to restore consistency to the rule base, since his or her identity now contradicts one of the experiences in the foundation. The second, however, is to do so in a way that retains decidability and coherence while accommodating the new belief.

Given the need to address problems of identity, individuals will have open to them a variety of strategies for searching through past experiences for ideas. Self-consciousness implies that all such strategies will be represented in the current rule base. However, one general presupposition that can be made about such strategies is that none will involve gratuitous deliberation, deliberation that is unnecessary
A General Model of Preference and Belief Formation

for attaining coherence. Another is that individuals will avoid making radical changes in the structure of their rule base unless it is necessary for them to do so.

Given this, Assumption 6: An individual facing undecidability or doubt will begin by adjusting orientations toward statements in the rule base without changing its parametric form. The individual will then consider new rules from the identity frame incrementally. At each step he or she will make the minimum adjustments necessary, and deliberation will stop once coherence has been attained.

In general, I assume that people avoid engaging in ideational innovation, the creation of new ideas that do not exist even latently in their identity frames. There are a number of reasons for this. The first is that uninhibited ability to innovate removes the parametric constraints from an individual’s preferences and beliefs and therefore makes the problem of rule base choice undefined and somewhat paradoxically more difficult, just as the lack of constraints imposed by a rule base somewhat paradoxically make it more difficult to optimize. Second, there are usually reasonable grounds for assuming a sufficient variety of ideas available in a cultural milieu to permit most people to attain coherence via orientations on those ideas. The point of the model is to find out through which preferences and beliefs individuals ultimately achieve coherence, not if. Indeed, the overproliferation of ideas that could conceivably achieve coherence can be as much of a problem as their shortage, particularly when expected coordination of ideas within a group is a major factor in perceived optimality. Third, when more ideas are needed there will be no shortage of ideological entrepreneurs anxious to generate new frameworks, often reflecting their own self-interest.

Even so, there will be certain types of self-generated ideas that are readily available to all individuals within the rule base itself. In particular, these will involve transforming the constants and wild cards of existing rules in their identities, without fundamentally changing their parametric form. While the rules in an individual’s foundation do not specify the disposition of any contingency like the rules in any identity, foundational rules nonetheless remain in a rule base as a necessary means for parsing new ideas, and for rejecting those that have no meaning vis-à-vis the assignment of wild cards and constants to classes. Nonetheless, they provide another resource, which is to provide ready-made alternatives to rules in an identity. In particular, by grouping constants into classes represented by a wild card, they provide alternatives to any rule in which a wild card or a constant is represented on the right-hand side of a statement. In particular, a foundation provides a way of transforming a rule without changing its parametric form by arbitrarily replacing constants or wild cards of a finite class with other constants or wild cards of that class, thus essentially treating such concepts as parameters.

Take any statement of the form $x = f\hat{y}$, where $f\hat{y}$ is an expression containing the wild card $\hat{y}$ over some class $Y$. Given this, $\hat{y}$ may be replaced by any constant $Y \in Y$, as long as the number of entities in
Y is finite. Likewise, if \( x = f y \) may be replaced by a \( y' \), which is a wild card over \( Y' \subset Y \). This kind of transformation will be called a specification, limiting the scope of an idea to a particular instance or subset of instances.

On the other hand, any statement of the form \( x = f Y \) can be transformed by replacing constant \( Y \) with a wild card \( y' \in Y' \subset Y \) or another constant \( Y' \). The former can be called generalization, opening the scope conditions to make them universal, and the last replacement, modifying the conditions without expanding or contracting them. More selective types of transformations will also be possible, as long as they require only the replacement of one finite subset of class by another. One thing to note is that constants are simply wild cards ranging over a class subset of a single member. Hence, we can treat them as part of the more general issue of wild card transformation.

Not only is such transformation possible for statements in a rule base, but it is also possible for any new idea that an individual may consider. Thus, even if an individual is forced to draw upon his or her identity frame, this ability alone can be enough to transform and provide diametrically opposed alternatives to culturally hegemonic ideas that may be the only ones available, turning an idea into its antithesis prior to adoption. At the same time, it does nothing to change the structure of the idea, and hence allows ideas in a cultural milieu to set the parametric form of an individual’s rules even when they fail to shape the dispositions he or she assigns to parameters.

However, transformation will take place only over finite classes of entities. While optimization over parameters on infinite classes of numbers is simply a constrained optimization problem that can be addressed by techniques like linear and nonlinear programming, classical optimization theory, and so on, self-definition implies that all such methods employed by an individual will be included as statements in a rule base. In other words, the solution offered by a constrained optimization technique must be represented in a rule as the disposition the contingency being optimized, and the steps involved must be specified as a recursive function on the right-hand side of the rule. Absent any algorithm whatsoever, the only recourse for an individual will be random search within the space of alternative dispositions. However, random search within an infinite space is by definition not guaranteed to halt and hence is not decidable.

Another way to look at this is to view each technique as a “solution generator” over an infinite class. Given a choice of techniques for generating solutions, an individual may assign probabilities to the chance that each one will generate the correct solution. Given this, the class of alternatives that he or she is optimizing over will not be the infinite class but the class of solution generators. Hence, an infinite class can be dealt with within a choice problem but only by considering a restricted subset of its possible dispositions.

The assumption implies that individuals optimize over doubt, but do so in a way that avoids gratuitous deliberation. If necessary, they will adopt and adjust their orientations toward new ideas in their identity frame, but they will maintain the rule base as close as possible
to the status quo in form. This view of optimization appropriates the notion of problemistic search, one usually associated with bounded rationality models rather than optimizing ones. The reason this is possible is that doubt has a universal optimum of zero. Individuals can always tell when they are in a state of coherence (i.e., peace of mind), so there is no need to specify an aspiration level for each choice and individual or for an individual to consider all alternatives before knowing what is optimal. They do not optimize over deliberation time, since this leads to infinite regress and is impossible, but they avoid it once their identity needs are fulfilled.

Some clarifications can be made about the way in which doubt minimization should be interpreted. First, while individuals may adopt certain beliefs over others without an informational basis for doing so, this does not imply that they are engaged in a process of self-delusion. Instead, it reflects the notion that, in order to engage in optimization, individuals need to tentatively adopt certain working beliefs that allow them to calculate expected utility and that they will do so in a way that is consistent with information that they possess from observation (assumption 4) but best preserves a sense of coherent identity. Unlike beliefs based on information, such beliefs will be defeasible in the legal/philosophical sense, i.e., they may be revoked or modified if contrary information arises or they no longer serve to preserve coherence.

Likewise, the notion that individuals consciously adopt preferences does not imply that they are able to “choose their affect.” Preferences are viewed here as consciously held values rather than indicators of emotional states and define parameters for optimization independent of whether it is directly pleasurable in a hedonic sense to attain them. At the same time, however, rational individuals will not ignore the influence of affect on their ability to carry out actions and must take this into account in choosing preferences, something that is captured quite crudely by assumption 3.¹⁰

Finally, the idea that individuals minimize cumulative doubt attached to past actions as well as future ones does not imply that individuals attempt to view their past selves as infallible or eliminate all perceptions of past mistakes. Indeed, individuals who change their preferences and beliefs will be aware that they have changed them and therefore will necessarily view their past attitudes as mistaken. Moreover, they will not compare their strategies to all available after the fact, but only those that were part of a feasible strategy before the fact, i.e., were optimal under some alternative set of preferences and beliefs available in their identity frame. Coherence is not the same thing as perceived perfection, but rather the ability to view one’s actions teleologically, as fulfilling some coherent purpose, regardless of whether one was aware of it at the time.

¹⁰ However, it is admittedly somewhat of a kludge, and there is a discussion in chapter 7 of deeper assumptions about the relationship between preferences and affect as well as a more radical separation between attitudes and self.
Basic Implications of the Model

Two questions that arise, however, are how individuals will know whether their choice problems are decidable or not, or how they will modify current ideas and search for new ideas to adopt in a way that does not lead to infinite deliberation.

As mentioned, not all problems have solutions, and certain classes of problems that do will not yield them in a finite amount of time, even given arbitrarily high levels of computing power. Hence, decidability is not a function of the boundedness of an individual’s rationality, but rather the nature of the problem itself. The class of problems that are guaranteed to yield finite-time solutions is that for which solutions can be specified recursively, i.e., through application of a series of functions in which the result of one function fully determines the next. \(^{11}\) Recursive sets are usually seen as equivalent to the informal notion of computability, the existence of a computer algorithm that is guaranteed to halt and generate the correct answer. \(^{12}\)

While there is no general way to specify the form of decidable as opposed to undecidable problems, proofs of decidability often revolve around finding an appropriate metric through which it can be shown (or shown not to hold) that an algorithm cuts down by some minimum amount the size of the problem for each step. One way to do so is to view a formal theory as a language, with its symbolic system and interpretations as syntax and semantics, respectively. As the usage implies, syntax is typically defined broadly as encompassing both grammar and logic, and semantics as encompassing both meanings and posited realities. A rule base can be viewed in such a fashion, with its foundation as syntax and identity as semantics. Given this, we can show the following:

Theorem 1: For a rule base that is finite, consistent, and in causal form, the following will all be decidable: (1) the well-formedness of any new idea vis-à-vis syntactic rules contained in an individual’s foundation, (2) the semantic consistency of any new idea with those in an identity, and (3) the possible dispositions for the utility attached to any specific action under any disposition of fates.

Proof of this theorem as well as others in this chapter, can be found at the end of this chapter.

Therefore, restricting the structure of a rule base to causal form can ensure decidability, regardless of the types of experiences and actions for which individual has to account. Such a result may explain why causality, despite being an ill-defined concept metaphysically, appears to be an inescapable part of any individual’s comprehension of reality. The crucial issue in whether a conceptual framework incorporating causality is useful is not whether any verifiable propellant force emerging from the right-hand side is responsible for the left-hand side but that order is consistent and transitive and that the specified relationship squares with experience. Other than that an individual

\(^{11}\) Recursive function theories derives in large part from Kleene 1936.

\(^{12}\) For a review of these issues with regards to general classes of computing problems, see Hartmanis and Hopcroft 1968.
need not have a clear view of whether the relationship in question is causal or constitutive (i.e., definitional) or whether it is empirical or normative.

Indeed, for quasi-normative contingencies (e.g., justice) or even apparently empirical contingencies with unambiguous normative valence (e.g., terrorism), the distinction between a causal idea and a definition is often quite nebulous, yet people often use such concepts to guide them in everyday life. Hence, while both causal ideas and definitions are assumed to be contained in an individual’s rule set, there is no assumption that he or she has a complete categorization of which is which, and such categorizations per se will not influence the effect of such ideas on choices.

One thing that the theorem does imply, however, is that individuals gain from having a common conceptualization of causal order, regardless of what it is based on. While a rule base that is finite, consistent, and in causal form can evaluate any new statement for wellformedness and consistency vis-à-vis its own rules, it cannot always render every available statement into usable causal form. However, sharing a common causal order ensures that such general parsability exists.

The theorem is also relevant to understanding why debates over the definitions of words are often as heated as debates over empirical or normative ideas. The idea $x = fy$ can serve as a rule for $x$, and can thus serve as a potential candidate for preference and belief change and potential guide for action, regardless of what the “=” stands for metaphysically. In any of these cases, an individual will consult the her beliefs about the contingencies in $y$, transformed by operator $f$, in order to determine her beliefs about $x$. Indeed, to the extent that the most prevalent ideas about a particular contingency are typically definitional ones rather than causal ones, it is likely that such rules will be the ones that individuals take as foundational rather than part of his or her identity. Indeed, while an empirical or normative idea may be rejected as ill-formed because its claimed relationship between contingencies clashes with dictionary definitions of those contingencies, the latter is rarely the case. Hence a definition is typically the most powerful of ideas in shaping rules and overturning a definition can have a more powerful impact on individual actions than overturning an empirical or normative idea.

Before continuing on to address theorems about preferences and beliefs, one thing can be noted:

**Proposition:** Individuals will adopt as a default assumption the idea that their future selves and all other individuals will be rational optimizers and that they will more generally act according to the assumptions 1 through 6.

This follows from self-definition and self-referentiality. Given that individuals act and think according to assumptions 1 through 6, these assumptions will be in their foundation. Not only will an individual be aware of these assumptions, but he or she will be able to extend the assumptions by generalizing from constants (the individual herself at
the current point in time) to wild cards (all individuals at all points in
time). Hence, rules about $P_{i,j}$ can be extended to all $P_{i,j}$.

Moreover, since the individual will be able to engage in this
generalization as soon as he or she attains the ability to engage in
rational deliberation, the generalization will form status quo beliefs
regarding other individuals, and these will be retained until they prove
inadequate for addressing coherence problems related to such beliefs.
However, such refutation is unlikely due to the inherent flexibility of
this induced common knowledge assumption. Just as they can decide
among preferences and beliefs to address coherence problems related
to their own actions, individuals can ascribe different preferences and
beliefs to other individuals in order to explain their behavior. Hence,
rationality itself is less of a falsifiable model of others than a way
of interpreting their patterns of behavior. It is not the assumption
of rationality per se that allows an individual to predict another’s
behavior, but rather knowledge of the way rationality will be shaped
into a set of rules under a particular cultural background and set of
experiences.

Common knowledge of this kind is important because, in
conjunction with cultural salience, it allows us to analyze ideas as
an arena of strategic interaction. The assumption is analogous to
the common knowledge of rationality in conventional rational choice,
though it focuses on attitudes rather than actions and explicitly extends
common knowledge to future selves as well as others (something that
is implicit in the conventional common knowledge assumption). Like
common knowledge of rationality, it allows individuals to iteratively
narrow the range of potentially optimal attitudes by predicting which
attitudes might be optimal for other individuals. On the other hand,
the implications of this for choice of preferences and beliefs as well as
for action, are qualitatively different and stronger than the implications
of iterated dominance procedures for action. Individuals will always
form beliefs that eliminate strategic uncertainty, and they will be shown
to often apply a quasi-Kantian perspective to their preferences.

Preference and Belief Change
This section will focus on ways in which doubt generated by an action
can be reduced by changes in preferences and beliefs. A number of
rather basic theorems will be provided, each linked to empirical findings
or theoretical issues in social science research.

Theorem 2: Doubt for an action will be reduced by increasing the
causal weight between any good that is deterministically linked to a
chosen action and utility or any other good that is on its causal path
to utility.

Intuitively, increasing one’s valuation of any good that is linked
to a chosen action will increase the utility it provides vis-à-vis other
actions in all states of the environment. However, this change can be
direct or indirect. It can involve changing the causal weighting of the
good within a utility function itself or changing its weighting vis-à-
vis some other good that is on its causal path to utility. While the
first can be seen as changes in preferences and the latter as changes in
beliefs, in either case it creates a stronger linkage overall between a good and utility. Indeed, belief change affects doubt because it affects the perceived instrumental (extrinsic) utility attached to particular actions. As such, while constraints on change may be different, the process by which doubt is reduced is fundamentally the same.

The theorem can be used to analyze a number of phenomena in social psychology. For instance, the literature on postdecisional preference change has posited that individuals will tend to increase their relative liking for a particular good once they have chosen it from among a set of alternatives.\footnote{Brehm 1956.} This phenomenon is related to one in economics known as the endowment effect.\footnote{Thaler 1980, 43-47. The endowment effect is actually broader, and relates to changes in valuation that are triggered by the mere possession of a good, regardless of whether possession is due to an individual’s actions. One therefore needs to separate the extent to which the effect is caused by motivated preference change and the extent to which it is caused by nonlinear utility functions and/or nonexpected utility.} These findings coincide with the implications of the coherence model since possession of the good is causally linked to the action taken. The model furthermore allows us to link preference change to actions even when individuals have no reason to believe that the actions were irrational. However, it would also imply that such preference change will depend on the existence of a perceived risk that the choice an individual has made will not turn out in the end to be best available. Where clear differences exist in the monetary value of different goods, the model would predict that no preference change would occur.

It can also be used to analyze the phenomenon dubbed the functional autonomy of motives,\footnote{Allport 1937.} in which actions chosen to achieve certain goals eventually come to be valued in and of themselves and are sometimes continued even when their connection with the original goals is severed. The paradigmatic cases of this are the employee who is diligent beyond the point at which any link exists between additional work and income, or the weekend athlete who exercises to such an extreme extent that it actually harms his or her health. This phenomenon is connected, albeit indirectly, to the large body of work in behavioral psychology on the development of secondary reinforcers, arbitrary consequences of certain behaviors that trigger further behavior even when the rewards (primary reinforcers) they were originally linked to are removed.\footnote{For an unusually motivation-oriented take on secondary reinforcers, see Prelec and Herrnstein 1991, 333-35.}

Functional autonomy of motives can be thought of as one form of postdecisional preference change, albeit one that is usually associated with repeated action over an extended period. However, its focus is specifically on the intrinsic valuation of a type of action, and hence it has certain special qualities. Intrinsic valuation is most useful for reducing doubt when there are alternative actions that might generate
comparable levels of the goods provided by the action in question. If alternatives did not exist or were clearly inferior, individuals could raise the subjective value of these goods in order to justify their choices and any sacrifices they are making. If an employee’s job is unquestionably the best way to make money, an employee can eliminate doubt over vocational choice by simply valuing money more. If sporting activity is unquestionably the one best way to improve health, a weekend athlete can eliminate doubt over his or her decision to engage in the activity by valuing health more. In either case, the behavior will not persist once the connection to extrinsic rewards is ended.

**Proposition:** Functional autonomy of motives will occur when individuals have doubts about the commitments they have made to an activity, and there are alternative activities that provide the same kinds of benefits as those for which the activity was originally undertaken.

Besides being postdecisional, doubt can also be predecisional. Furthermore, it may exist when consequences are unknown as well as when they are known. This allows the model to be applied to the contradictory phenomena Elster calls “sour grapes,” and “the forbidden fruit is sweet.”

Sour grapes occurs when individuals change their preferences in a way that increases their expected utility under current circumstances, while forbidden fruit occurs when change is in the opposite direction. Neither sour grapes nor forbidden fruit is presented as, or could be, the basis for a general model of preference formation. In a sour grapes model, individuals can generate arbitrarily high amounts of expected utility by simply incorporating contingencies that have unchanging positive values (e.g., the existence of the universe) into their utility functions with arbitrarily high positive weights. In a forbidden fruit model, they could similarly generate arbitrary low amounts, though it is not clear why they would want to do so.

While noting that both phenomena can be observed as part of human behavior, Elster does not attempt to specify the conditions under which each will happen. Elsewhere, he implies that social science is unable to provide useful general models of preference and belief formation and recommends the instead the identification of preference and belief mechanisms that may or may not occur.

While it is certainly true that no model will eliminate all significant indeterminacies and anomalies, it is certainly possible to address the types of preference change Elster describes. The coherence model implies that the key factors in preference change are not so much an individual’s circumstances, but as the perceived effect of his or her actions on these circumstances relative to the perceived effect of other actions available. More concretely, it implies that if an individual has rejected or plans to reject actions that could alter particular aspects of his or her existence, he or she can reduce doubt for this choice by viewing alternatives as sour grapes and lowering their desirability. If, on the other hand, he or she has chosen to reject the status quo and pursue a change, the individual can reduce doubt by viewing outcomes

---

17 Elster 1983c.
18 Elster 1989b, part I and Political Psychology, 5.
associated with change as forbidden fruit and raising their perceived desirability, even though such outcomes have yet to be realized and the actions have not yet been performed. As was discussed earlier, the main difference between past and future actions is that preference change linked to future actions will be transient if an individual changes his or her mind at some future point in time. However, in order for wishful thinking or forbidden fruit to reduce doubt, doubt must be linked to the unknown effects of change, not the probability that attempts to make changes will not be successful.

**Proposition:** Sour grapes effects will be associated with actions that reject any attempt to change the status quo. Forbidden fruit effects will be associated with attempts to change the status quo. Both will be associated with individual concerns and worries about the benefits that will arise once the status quo is overcome, not with concerns about whether the status quo can be changed.

This analysis suggests a general class of sources for the origins of preferences and can serve as a starting point for midrange theories that incorporate endogenous preferences formation. It instructs the theorist to look for choice sets in which the relative provision of material welfare is not higher with certainty for the rationally chosen action compared to other actions, then to look for goods whose relative provision is always higher for the action. It predicts that individuals will tend to increase their valuation of these items, even though they are not the part of the objective for which the actions were originally undertaken. Likewise, individuals will tend to decrease their valuation of goods whose relative provision for the action is always lower than that for other actions.

Individuals may reduce doubt through changes in subjective probabilities over rules as well as changes in causal linkages:

**Theorem 3:** Doubt for an action will be reduced by increasing the subjective probability of all dispositions of fate in which the action provides high levels of desired goods relative to other actions and by decreasing the probability of other dispositions proportionately.

Again, this is fairly intuitive. Choice of an action implies the existence of states in which the action provides higher amounts of certain goods than other actions do. Raising the subjective probability of these states and proportionately lowering the probability of other states will thus reduce existing doubt.

Doubt-induced subjective probability change can be applied to another type of postdecisional phenomenon, sometimes known as bolstering, in which individuals adopt beliefs that increase the perceived rewards that will be generated by the choices they have made, particularly when such choices involve greater gambles than the choice(s) they have passed up. Such changes are equivalent to decreases in the subjective probability of states in which the choice they made provides lower levels of benefits than other choices and proportional increases in the subjective probability of other states. As with postdecisional preference change, the model implies that bolstering takes place only when there is a perceived risk that the

---

choice will not turn out to be the best available. It also constrains such changes to be consistent with any evidence that individuals had observed.

The theorem can also be used to analyze and contextualize the illusion of control phenomenon, in which individuals have been found to systematically exaggerate the influence their choices have over important outcomes. Psychologists have often asserted that such exaggerations are often beneficial or even essential to psychological well-being, and they argue that this is one reason they may be adopted. However, the phenomenon seems to contradict another phenomenon known as learned helplessness in which individuals who have been unsuccessful in past efforts continue to perceive an inability to succeed, even when they are confronted with seemingly simple problems. It would thus be useful to specify when individuals will exaggerate control, and when they will exaggerate helplessness.

The model suggests that an increased perception of control is useful whenever an individual has chosen to pursue an endeavor aimed at a particular outcome, but perceives some possibility that the extent to which the outcome is realized may be insufficient to justify her choice. In such a case, an individual can reduce doubt by increasing the perceived influence of her actions on the outcome, since this will reduce the subjective probability that the utility provided by pursuit of the endeavor will be less than that provided by alternatives. On the other hand, if she has forgone pursuing the endeavor, her doubt will be caused by the perceived possibility that she might have been successful beyond her expectations. She therefore can reduce doubt by decreasing her perceived abilities in the endeavor. Barring observable information forcing her to believe otherwise, these belief changes can then influence future actions. This also indicates, however, that perceptions of control or helplessness will be connected to particular endeavor and the goals associated with them, and will generalize only if individuals are faced with similar choice problems in different contexts. It is thus possible for individuals to have an illusion of control in one area and learned helplessness in another.

**Proposition:** The illusion of control will be associated with decisions to participate in a certain endeavor, even though the increase in benefits provided by participation is uncertain. Learned helplessness will be associated with decisions to withdraw from an endeavor, even when there is some possibility that it might have paid off. In each case, belief change will be linked to concerns over the realizability of outcomes, not to concerns about the benefits of those outcomes.

The theorem is also relevant for analyzing the “wishful thinking” phenomenon, which has been discussed extensively by Elster, in

---

20 Langer 1975.
21 Taylor 1983.
22 For a survey, see Maier and Seligman 1976.
A General Model of Preference and Belief Formation

which individuals raise their subjective likelihood of favorable outcomes, hence raising their expected utility. Like sour grapes and forbidden fruit, wishful thinking has a counterpart in unwishful thinking. Similarly, neither wishful nor unwishful thinking can be the basis for a general and accurate model of belief formation since neither provides any constraints on belief adjustment that would prevent the adoption of arbitrarily wishful or unwishful beliefs and hence arbitrarily high or low expected utility. Furthermore, though mutually contradictory, each phenomenon is observed to occur in human behavior.

This theorem can be used to analyze both wishful and unwishful thinking and to specify the conditions under which each will occur. As with the sour grapes/forbidden fruit preference change, the direction of belief change will depend on the actions an individual has taken or plans to take and how it compares to alternative actions that are available. Suppose an individual has chosen a subjectively risky action over a safe one. It is in the nature of risky actions that they depend on favorable circumstances to pay off, whereas safe actions provide smaller but more reliable payoffs that vary little according to circumstance. If the payoff of the safe action is higher than that of the risky action under unfavorable circumstances, such an individual can reduce any doubt over the risky action by engaging in wishful thinking, decreasing the perceived likelihood of unfavorable circumstances and proportionately increasing the perceived likelihood of favorable circumstances. Likewise, if an individual has chosen a safe route over riskier ones because of fear that disaster will strike, he or she can reduce doubt by engaging in unwishful thinking.

Proposition: Wishful thinking will be associated with decisions to take a subjectively risky course of action when safer ones were perceived to be available. Unwishful thinking will be associated with the opposite type of decision.

Rewards and Preference and Belief Change
Besides looking at the circumstances that trigger different kinds of preference and belief change, it is possible to examine magnitudes of change. One fairly straightforward way of doing this is to examine the effects of perceived levels of existing incentives on the amount of preference change.

Theorem 4: The magnitude of change in causal weighting vis-à-vis utility of a good that is linked deterministically to an action will vary inversely with the relative amount of other desired goods provided by it. For prospective actions, this will hold only to the point at which the action is no longer optimal.

The main intuition here is that an action that provides higher levels of desired goods across states will generate lower levels of doubt if the action is chosen since the gap between its utility and that of other actions will be smaller in those states in which it does not provide the maximum utility and the number of such states will tend to be smaller as well. This in turn will decrease the amount of change in the utility

23 Elster 1983a.
weighting of a good linked to it that is necessary to eliminate doubt and restore coherence.

This theorem relates to a number of findings central to dissonance theory in psychology. One such finding comes out of classic work on “forced compliance,” which showed that a small monetary reward given for performing a particular task led individuals to evaluate the perceived enjoyability of the task more favorably than a large reward.24 This literature is closely linked to studies on “overjustification,” which have shown that providing large extrinsic incentives for an activity can reduce the amount to which individuals acquire intrinsic valuation of the activity.25

These types of findings describe situations in which the retrospective doubt produced by an action is shifted up or down by manipulating the observed effects of the action on the provision of some good known to be valued by all individuals involved. The theorem would thus predict that manipulations that decreased the utility of the action would increase the amount of positive change that occurs in the utility weighting of another good linked to the action, since more change would be necessary to restore coherence. Unlike these findings, however, it would extend preference change to goods that are believed to be causally linked to the action as well as to those constituted by it (i.e., intrinsic valuation of the action). It would also imply that if multiple goods are constituted and/or causally linked to an action, the one chosen for preference change will be highly dependent on the salience of a particular idea at the time adjustment occurs.26

The theorem also has relevance for findings on effort justification that have shown that high costs for entry, such as costly initiation, can increase the value placed on activities within a particular organization or group.27 In this case, doubt is shifted through manipulation of extrinsic costs (the difficulty of initiation) rather than extrinsic benefits, but the results are similar, though in the opposite direction. The theorem highlights the essential mirror-image nature of the induced compliance and effort justification effects since decreases in benefits and equal increases in costs have equivalent effects on doubt and hence on preference change.28

24 Festinger and Carlsmith 1959.
26 For instance, the early forced compliance experiments asked subjects to specify the level of enjoyment they gained from a particular activity. This in turn made the intrinsic valuation of the activity a salient idea available to subjects. The theorem would imply that if other doubt-relevant ideas had been presented as well, e.g., through a question about the value subjects placed on the advancement of science, there would have been a positive statistical relationship between orientations toward them and the low-reward group as well. Moreover, one would expect an inverse relationship between the presence of such a question and the amount of intrinsic valuation subject acquired for the activity.
28 In economics, Thaler refers to this as the sunk cost effect. See Thaler 1980, 47-50.
Looking at preference change from a choice-theoretic perspective allows us to reframe the long-running controversy between dissonance and reinforcement theorists. The former argue that an inverse relationship exists between rewards for an activity and its later persistence, and the latter arguing that a positive relationship exists. In particular, reinforcement theorists have shown that stronger reinforcement for a behavior increase the time required for extinction, i.e., discontinuation of the behavior once the reward has been withdrawn.\footnote{Skinner 1953, 69-72.} However, findings of dissonance experiments suggest that larger extrinsic incentives lead to less intrinsic valuation of a behavior, which in turn would reduce its persistence after the rewards are withdrawn.

Forced compliance, overjustification, and effort justification typify the dissonance literature in being concerned with retrospective preference change (i.e., they consider situations in which the relevant action has already been committed). Under such a condition, the model would predict that preference change will always be in the same direction, and will increase when incentives for the action decrease. On the other hand, the reinforcement literature deals with conditions under which choices are repeated and a behavior may or may not occur for each opportunity. Furthermore, by the very nature of their approach, behaviorists cannot measure the strength of a reinforcer except by measuring the frequency of behavior it elicits. Hence, weaker reinforcers will be those that are insufficient to elicit the maximum feasible frequency of response. Under such conditions, intrinsic valuation of the behavior will increase the doubt generated for periods in which the behavior was available but not chosen, and thereby make it impossible to eliminate cumulative doubt for all periods. As the reinforcer becomes stronger, the effectiveness of intrinsic valuation increases, up to the point at which it is strong enough to elicit the maximum feasible frequency. Any strength of reinforcement beyond this will be impossible to measure, and hence overjustification effects cannot be observed.

The theorem is clear that the relationship between rewards and the intrinsic valuation of an action, and hence the tendency to persist in that action, is neither monotonically positive nor monotonically negative when individuals have a choice of whether or not to perform the action. It would suggest that the absolute magnitude of preference change necessary to eliminate doubt for a chosen action would be greatest on the boundary where rewards are just high enough to make an individual indifferent between performing the rewarded action and not performing it. On the other hand, the sign of the preference change would be different on either side of the boundary, leading to a large chasm in between.

The relationship between extrinsic incentives and preference change in the case of dichotomous choice looks (very) roughly like figure 1, given convex and monotonically increasing linkages on the causal path between a good and utility. This suggests that attempts to induce an individual to maximally value such an action by shifting...
Choosing an Identity

overall reward levels are a very tricky business. Preference change will be largest when rewards are just great enough to elicit the desired behavior, but a slight underprovision of incentives will cause the individual to reject the behavior and will lead to an equally large change in the opposite direction. Unless the would-be preference inducer knows the subjects’ preferences precisely, any attempt to maximize preference change in this fashion may backfire.

Fig. 1. Relationship Between Rewards and Intrinsic Valuation

The theorem can be extended to discussions of reactance, the phenomenon in which externally imposed barriers to an individual’s ability to obtain a particular good lead to increased efforts to obtain it. Reactance theory has explained such actions as the result of an individual’s inborn desire to preserve his or her own freedom of action. The coherence model of course does not assume a preference for freedom a priori, but it can be used to predict the circumstances under which

30 Brehm 1966.
reactance might occur and when it will be associated with acquisition of such a preference. As long as barriers imposed from outside do not change incentives sufficient to prevent continued pursuit of a good, an individual will continue to pursue it or may even increase his or her efforts.

However, more than one explanation can be provided for this. First, even in the absence of preference change, barriers influence effort only to the extent that they influence marginal returns to effort, not total returns. In other words, restricting an individual from having a particular good may lower his or her provision of the good, but it may not lower the increase in the amount of the good that results from a fixed amount of effort to obtain it. In the most extreme case, a total lack of restriction will inspire no effort, since no effort will be required to obtain the amount of the good the individual desires. At the other extreme, however, barriers that are clearly impenetrable will also inspire no effort, since effort will not pay off in any improved provision of the good. This implies that the highest levels of effort will be found in the middle range, where barriers are expected to be strong enough to require effort on the part of an individual to overcome but not strong enough to prevent her from obtaining the good or to make effort suboptimal, and where there is some lack of certainty about the precise costs involved in overcoming the barriers.

It is also in this middle range that we would expect preference change, particularly when the barriers are expected to be just barely weak enough to make effort optimal. However, the nature of this preference change will also depend on the nature and context of the choice. If the barriers an individual is analyzing pertain to a specific set of goods being pursued, he or she will not add contingencies to the utility function, but rather adjust orientations within it by increasing valuation of these goods.

One reason an individual would acquire an intrinsic valuation for freedom would be if he or she expects to barriers to be placed upon her actions in future periods, but is unsure about the kinds of goods involved. Such a situation will exist when an individual expects to be under a diffuse kind of coercive supervision extending across a full range of activities and expects control to be troublesome but ultimately ineffective. Under these conditions, coherence cannot be achieved by adjusting valuations for particular goods, and hence an individual may adopt preferences relating to personal freedom. When supervision is expected to be undertaken by a particular individual or institution, however, he or she may also conceptualize these preferences as hostility to or desire to be independent from the particular individual or institution in question, depending on the other types of choices faced and types of ideas that are salient when he or she is deliberating.

Reactance theory must contend with its own opposing mechanism, the “Stockholm Syndrome,” whereby hostages acquire great loyalty to their captors over time. One can explain this syndrome by examining it within individual choices about whether or not to resist capture or escape. Hostages who remain in captivity and obey orders but are unsure about whether this is the right course can reduce doubt by
acquiring preferences for outcomes that depend on their compliance to their captors. As with other forms of attitude change, examination from a choice-theoretic perspective allows us to contextualize these mechanisms within opposing sets of conditions.

Proposition: Reactance arises when significant but subjectively inadequate barriers of varying strength are imposed on an individual’s freedom of action. This will result in freedom-oriented preferences only if such barriers are diffuse rather than specific to particular goods.

Like rewards levels, reward volatility has a systematic effect on preference change:

Theorem 5: The magnitude of change in the utility weighting of a good that is deterministically linked to an action will vary positively with the volatility in the amount of other goods provided by the action relative to other actions.

Reward volatility influences preference change through its effect on the overall amount of utility at stake in the choice. For every state in which the chosen action fails to provide maximum utility, the gap between the amount of utility it provides and the maximum feasible amount will be larger. Hence doubt and the amount of preference change required to restore coherence will be greater as well.

While reward volatility has apparently received less empirical attention than reward levels in psychology, the theorem can be used to provide a motivational underpinning for seemingly paradoxical findings in reinforcement theory. These show that behavior that is intermittently and randomly reinforced tends to persist for a longer time after the reinforcement is eliminated than behavior that is reinforced continuously or along a fixed schedule. Since such variable-ratio reinforcement increases the subjective volatility of reinforcement provided by the behavior being elicited, the theorem suggests that it will increase the doubt linked to that behavior and the tendency to value the behavior itself, even if the average level of reinforcement it provides is the same as an alternative schedule based on fixed-ratio reinforcement. This in turn increases the probability that the behavior will persist in the future despite withdrawal of the reinforcement. This suggests that, as long as the expected utility of extrinsic incentives is sufficient to elicit a certain behavior, providing randomized and highly variable amounts of such incentives will increase the amount of intrinsic valuation of the behavior itself.

The theorem is relevant to a phenomenon noted by Hirschman: for routine activities whose results are predictable, the “separation of the whole process into means and ends, or costs and benefits, occurs almost spontaneously.” On the other hand, when costs and benefits are uncertain, undertakings take on their own value “as though in compensation for the uncertainty it is this striving that is endowed with the feeling of already having a pleasurable experience.” The model presented here makes it possible to understand why this compensation

31 Skinner 1969, 118.
32 Hirschman 1982, 89. For a later comment along similar lines, see Hirschman 1994, 344.
occurs. Greater amounts of doubt caused by uncertain outcomes lead to greater preference change, which makes an activity valuable for its own sake.

It is worthwhile to emphasize that doubt is increased by volatility in relative, not absolute, rewards. Hence, it is not linked solely to the risk associated with the chosen action, but rather to the riskiness of all actions in the choice set. For instance, an individual who is choosing between a lottery and a certain outcome with equal expected utility will experience identical doubt, regardless of which choice is made. If he or she chooses the certain outcome, doubt will be caused by the possibility that he or she might have won the lottery. Hence, individuals cannot avoid doubt by avoiding gambles; it is awareness that gambles are available that causes doubt.

One important difference between the effects of reward level and reward volatility is that increasing reward volatility will rarely lead to a reversal in the sign of preference change or a precipitous inversion of the amount of change. Increasing volatility without changing the expected level of reward for each action will have little influence on whether or not a particular action is optimal as long as individuals are fairly risk-neutral over the goods in question. Moreover, risk aversion or risk seeking can be partly compensated for by distributing volatility across all actions in the choice set.

This means that engineering attitude change by manipulating reward volatility is much easier than doing so by manipulating levels. The theorem implies that if one can place individuals in situations where they encounter opportunities with observable high risks but high rewards as well, and that are unanticipated and dissimilar to opportunities that an individual has already experienced or anticipated, one can typically induce large amounts of attitude change, in reaction to either taking the opportunities or refusing them. Even then, of course, the types of goods for which a preference is acquired will depend on the ability of an individual’s current attitudes to eliminate doubt through adjustment and the types of ideas that are salient to the individual at the time deliberation occurs.

Implications for Dynamic Choice
The previous section focused on optimization of preferences and beliefs over strategy choices without explicit analysis of the dynamic context in which choices take place. However, consideration of strategies executed over a long period of time will force an individual to consider both the effects of his or her actions on future selves and the types of actions he or she can expect future selves to carry out.

Utility Discounting
One important factor in dynamic choice is the way in which individuals will discount their future utilities. No assumptions have been made about utility discounting in this model. Indeed, it is one aspect of the overall choice an individual can make about her preferences. However, the logic of the coherence model allows us to draw fairly
strong conclusions about the types of discounting functions individuals will adopt. In particular:

**Proposition:** Individuals will adopt preferences and beliefs that are time-consistent.

Self-referentiality implies that an individual actually believes she will carry out her intended actions. Indeed, such a notion is inherent to the rational optimization model. A rationally optimizing individual will choose actions to maximize expected utility and, by common knowledge, will assume that future selves will do so as well. A time-inconsistent set of beliefs is contradictory to the assumption of rational optimization, even if not all beliefs are information-based. If you expect to believe something that you do not currently think is true, self-consciousness implies that you believe that some new experience will come your way that will make the belief efficacious in eliminating a contradiction in your rule set between your foundation and identity (e.g., an anomaly vis-à-vis your identity). However, if you believe that such an experience will happen, you will experience doubt unless you hold the belief currently.

A time-inconsistent set of preferences is equally problematic because the only way to retain logical consistency and self-referentiality is to have intentions toward your future actions that are not optimal given your current utility function. There is no way to get around this by adopting an algorithm that causes you to perturb your utility function at some regular interval, since self-definition implies that the algorithm itself will be part of your beliefs, and self-consciousness implies you must therefore anticipate any perturbations that will occur. Given this, you must expect that you will suboptimize given your current preferences, which in turn will generate doubt. Thus, in order for her intended actions to maximize expected utility for present and all future selves, an must set her utility function to be time-consistent in ordering sequences of outcomes for all present and future selves.

One important corollary that arises from this, is that, from the beginning, a coherent identity will require that an individual value future utility as well as present utility. Therefore, although assumption 4 posits an initial utility function which covers only short-term material welfare, this proposition implies that an individual’s concern for his or her material welfare must immediately be spread to the foreseeable future in order to maintain coherence. Discounting rates may be arbitrarily steep, but valuation of the future must be nonzero in order for the choice of future actions to be decidable.

Rational choice theories of choice over time typically specify a utility function as a discounted sum of transient utilities attached to particular periods. The whole notion of discounting implicitly reflects an assumption that utility is determined by a sequence of consumption values rather than timeless existences. Furthermore, attaching the label “utility” to these consumption values as well as to their cumulation implies that cumulation is additive. Given these implicit assumptions, the only way to maintain time consistency is via geometric discounting, a property that makes geometric discounting attractive to rational choice theorists. If we abandon the assumption of additive cumulation
across periods, there are a much wider range of time-consistent discounting functions that can be considered, even if we accept the rather limited hypothesis that the only contingencies relevant for utility are levels of consumption. For instance, an individual may want to distribute consumption value as evenly as possible across periods, may want to cycle between periods of high consumption and those of low consumption, and so forth.

The proposition does not imply, however, that an individual will always adopt time-consistent discounting regarding past periods as well as future ones. Since all actions that affect past periods have already been committed, he or she can conceivably adopt discounting functions that reorder the utility attached to different past actions without affecting her intentions. It is for a similar reason that conventional rational choice theories can get away with assuming that an individual places zero value on past transient utilities.

Because time-inconsistency regarding the past does not inherently generate doubt, it is feasible to disown the past, but not feasible to disown the future. Under situations in which coherence problems related to past actions cannot be addressed any other way, an individual can simply be “reborn” into a new identity. However, there are two necessary conditions for doing this. First, the past can be disowned only to the extent that past actions are not seen as having affected current utilities. If they have, then disowning your past self utilitywise does not allow you to wipe out these effects on your present self. Second, in order for rebirth to have a positive effect on coherence, it must be unanticipated and thus infrequent. An individual who adopts a strategy of repeatedly abandoning past selves will anticipate doing so and thus have no way of justifying current actions.

Substantively, the proposition coincides with a fairly intuitive notion of coherent identity. An individual cannot intend his or her values to change over time without introducing a sense of incoherence. In other words, coherence depends in some sense on viewing values as timeless normative truths rather than ephemeral adaptations to short-term doubts. An individual must view any future failure to act according to current values as irrational, barring unanticipated new experiences. This implies that weakness of will and the need to self-commit through binding oneself should properly be seen as outside of a coherent notion of rationality rather than as the product of a peculiar kind of utility function.

Besides the shape of a discounting function, we can draw conclusions about the conditions under which individuals will adopt higher or lower utility weightings in future periods. From a fairly straightforward application of the theorem on changes in causal linkages, it follows that individuals can reduce doubt by increasing their relative valuation of future consumption if the optimal action is causally linked to higher future consumption levels than those provided by other

---

33 In chapter 7, I will briefly argue that weakness of will and various nonexpected utility phenomena are a product of incomplete de-anchoring of utility and probability from affect and salience.
actions. Thus, if the optimal action, given current preferences, carries some risk of poorer short-term outcomes than other available actions but is certain to provide better long-term outcomes, then individuals will tend to increase the value they place on the latter as opposed to the former.

This implies that a willingness to defer gratification can be induced in children through socialization, but only under a fairly specific set of conditions. First, sufficient short-term incentives must be provided at the beginning of socialization so that children will choose to defer gratification, even if they have not already internalized a preference for doing so. Second, the magnitude of short-term incentives must eventually vary in unpredictable ways, being delivered at some times and not at others, even if the child engages in the deferring action. Finally, the long-term reward must be provided reliably as long as the child engages in the deferring action. As time goes on, the average level of the short-term incentives can be gradually reduced. Current experimental research seems to confirm that childhood socialization is a factor in the tendency to defer gratification, though it focuses more on a child’s ability to carry out deferral than on the desire to do so.\footnote{See the overview in Mischel, Shoda, and Rodriguez 1989, 147-64. More specifically, Michel et al. focus on the attention strategies through which a child can resist the temptation to take a short-term reward over a superior long-term reward. They survey children to make sure that the long-term reward is always preferred to the short-term alternative, so the desire to pursue the long-term reward is not an issue. While imperfect ability to carry out intentions cannot be analyzed within a pure optimizing model of action, chapter 7 discusses how wrestling with one’s affect and imperfect rationality can be used to account for attention strategies.}

**Proposition:** Increased willingness to defer gratification will be induced by choice sets in which the optimal action provides an uncertain level of short-term benefits relative to other actions but a reliably higher level of long-term benefits.

### Preference and Belief Change Over Time

Looking at it from a long-term view, we can try to analyze the types of preferences and beliefs that individuals will acquire over their lifetimes and how they will adjust to interactions between their actions at different points in time.

**Proposition:** Individuals over time will tend to shift their utility weightings to goods whose provision can be linked to their actions in a deterministic causal fashion.

Even if an individual starts off with a utility function that reflects only material welfare, there will be powerful forces compelling him or her to acquire different sorts of preferences. One characteristic of material welfare is that, while it encompasses several dimensions of an individual’s existence, endowments along each of these dimensions can be traded off against each other through barter, selling, and purchasing. The very fungibility of the dimensions of material welfare is what allows it to be approximated by a single indicator, wealth.

However, it also means that there will be for many individuals multiple alternative strategies that, according to their beliefs, could
potentially provide maximum levels of material welfare. Given this, individuals with purely materialistic preferences will face doubt regardless of which strategy they choose. One way they can address this is by shifting their preferences from material welfare to goods that are causally or constitutively linked to a particular strategy.

Such a phenomenon can be used to explain the process through which, over time, preferences for higher-order goods develop from a base of self-interested materialism. The precise good selected will depend on the nature of the choice set and the culturally salient ideas available to an individual at the time the choice is made. If the choice dilemma is related to vocation, for instance, doubt can be reduced by acquiring a sense of attachment to one’s profession or some activity that is associated with that profession and no other. These goods can only be generated by participation in the profession, hence valuing them raises the utility offered by the vocation compared to all others. Moreover, once a vocation is entered, sunk costs invested in it through past actions can be justified in a similar manner. Hence, repetition of an action in the presence of doubt, even if originally motivated by material preferences, would tend to lead to increased valuations over time for higher-order goods linked to that action.

This notion that materialism precedes the acquisition of higher preferences is consistent with Maslow’s well-known “hierarchy of needs,” which posits a progression from material to belongingness to self-actualization needs. Maslow’s theory, however, is premised on the assumption that higher needs will be pursued after only lower needs are satisfied to some adequate extent. The explanation for the development of these needs is quite different here since the coherence model does not posit an aspiration level or specify a hierarchy. Instead, it views the acquisition of higher preferences as a reaction to multiple choices with uncertain relative returns for obtaining material welfare.

In this respect, at least, the predictions of the model appear somewhat less anomalous than those of the hierarchy of needs. While the hierarchy of needs predicts an roughly inverse relationship between wealth and materialism, the empirical evidence suggests a less straightforward relationship. On the macro level, while long-term surveys have suggested a declining level of materialism in advanced industrialized societies, the early stages of industrialization have typically been associated with a perceived rise in materialism. At the micro level, there is no clear evidence that wealthy people are less materialistic than poor people or that material success over time tends to lessen an individual’s desire for accumulation. It thus seems that whatever relationship exists between wealth and materialism is due to something more than satiation with material welfare.

As with many types of actions, we can look at efforts to acquire money along two dimensions, level and type. Such actions will cause doubt if either the maximum net returns to effort are not reliably positive or the relative returns to different types of effort are not certain. In either case, individuals can reduce doubt by acquiring

---

35 Maslow 1970.
Choosing an Identity

nonmaterialistic preferences. This suggests that materialism is not linked to wealth per se but rather to two factors. The first is the extent to which an individual perceives opportunities that will reliably increase her wealth. The second is the extent to which there is a path for obtaining wealth that with relative subjective certainty is at least as effective as all other paths available.

Proposition: Materialistic values are associated with conditions in which net returns to efforts to gain wealth are reliably positive and largely independent of expected returns beyond the minimal net positive level and there is a clear-cut path that is reliably better than others for gaining wealth.

Implications for Collective Choice

While preferences and beliefs belong to individuals, the process by which they are generated is collective. There are a number of dimensions along which this collective process can be analyzed. One relates to the different ways in which identities can be acquired as a by-product of action, and the other to the way in which groups and societies can try to induce certain identities among their members.

Collective Action and Altruism

One important form of shared identity is group altruism, the mutual incorporation of some aspect of another individual’s welfare in one’s own utility function. A number of rational choice scholars have shown how altruistic preferences can have a major impact on the choices of individuals and on collective outcomes. In particular, altruism can provide incentives for individuals to engage in cooperative action in cases in which it would otherwise not occur. Conversely, the existence of negative altruism can prevent cooperation when it otherwise would be possible. It is clear that altruism can explain cooperation that is anomalous under conventional assumptions, but the lack of falsifiable models to explain the origin and levels of altruism has hampered applying it to real-world environments.

---

36 Altruism here is defined in the way economists define it, as a kind of preference. In the social psychology literature, altruism is more often defined as a kind of behavior, i.e., helping, not the motivations for such behavior. See Krebs and Miller 1985, 1-3.

37 Collard 1978, chap. 4; Moe 1980, chap. 5; Hardin 1982, chap. 7; Margolis 1982, chap. 5; Taylor 1987, chap. 5; Chong 1991, chap. 4; Monroe 1996, part I; Chai and Wildavsky 1994.

38 Margolis’s model of altruism has a major deductive component, since it posits a kind of “satiation” effect for both self-interested and altruistic behavior, but as noted in chapter 2, it does not seek to explain whom individuals will be altruistic toward nor the level of their altruism. See Margolis 1982. There also exist a number of genetic evolutionary models of how altruism may spread given an initial group of individuals with such preferences. However, they can only explain changes in patterns of preferences over a long period of history. See discussion in Elster’s Elster 1979, chap. I and chapter III, pp. 144-6. Recently, Boyd and Richerson have come up with a “conformist cultural transmission” model which addresses this problem, but it still cannot explain the origin of altruistic preferences or changes in an individual’s altruism patterns over her lifetime. See Boyd and Richerson 1990, 111-32.
The primary insights of the coherence model for altruism can be revealed by treating altruism as a contingency within a utility function. In that case, if an individual engages in action that has nonzero doubt but always provides a net increase in the provision of a particular public good to the members of a particular group, the individual can reduce doubt for that action by increasing his or her valuation of the provision of the public good. Conversely, if an individual engages in action that has a nonzero doubt but always provides a net decrease in the provision of a public good to the members of a particular group, the acting individual can reduce doubt for that action by decreasing his or her valuation of the group’s public good.

When the goods in question are positively valued, the former type of action can be called cooperative with regards to the affected group, while the latter type can be called conflictual with regard to the group. This implies that if an individual engages in cooperative collective action for which his or her net benefit for participation is not reliably positive but where participation provides reliably positive amounts of a good for other members of the cooperating group, he or she can reduce doubt by valuing their provision of that good. If the good is positively valued by the group that is being benefited, the preference generated can be seen as kind of altruism toward members of the group.

Such a situation, in fact, is quite typical in cooperative collective action. In most cases, participation has consequences both for the provision of a good that is shared among members of the group and for the amount of private benefits that are provided only to those who participate. The key insight of Olson was that in large groups, an individual’s direct benefit from his or her contribution to the shared good will not be large enough to make participation rational. Hence, cooperation depends on whether private benefits (selective incentives) provided only to participants outweigh the costs of participation. Selective incentives, however, depend on accurate monitoring of whether or not each individual has participated. Whenever monitoring is less than perfectly accurate, there will be nonzero probability that participation will be suboptimal, even when the expected utility of participation is positive. Because monitoring might fail and the individual may have gotten away with shirking, doubt will be generated.

On the other hand, individual participation will in most situations have a fairly certain positive net effect on the level of the shared good. Therefore, an individual experiencing doubt over participation can reduce it by increasing his or her valuation of the provision of the public good to the members of the group.

---

39 Olson 1965, chapters 1-2.

40 I use the term “public good” to refer to those that are shared within a group rather than those that must be shared because of their inherent technical characteristics. As Malkin and Wildavsky point out, there are very few if any goods that are inherently nonexcludable, although the costs of making them excludable may be quite high. By labelling certain goods “public goods,” analysts simply foreclose the possibility that they may be provided privately. For the purposes of this analysis, the fact that certain goods are being provided collectively is more important than why this is happening. See Malkin and Wildavsky 1992.
Choosing an Identity

The perceived size of the individual’s contribution to the provision of the public good will not be crucial to the doubt-reducing efficacy of altruism, as long as it is reliably positive. Hence, participation even in rather large groups can lead to the development of altruism. Moreover, once altruism develops, it can lead to future cooperative action even when the levels of selective incentives are lowered, since the individual will gain an inherent benefit from his or her contribution to the collective outcome. Cooperative action that is initially undertaken for self-interested reasons may eventually become rational at least in part for altruistic reasons.

In those cases in which an individual’s actions lead to a reliable decrease in the provision of a public good or a reliable increase in the provision of a public bad, i.e., something negatively valued by those whom it affects, the result of doubt-reduction can be seen as negative altruism toward members of the group affected by her action. Thus, just as cooperative behavior can promote future cooperation among a group of individuals, conflictual behavior can inhibit future cooperation and promote future conflict.

It is important to note that the types of altruism being discussed here do not require that altruistic individuals incorporate the entire utility function of others into their own utility functions. Instead, they will care about the provision of particular goods to other individuals. Concentrating on the provision of goods avoids having to assume that individuals have direct knowledge of others’ utility functions. It also allows a number of different types of “altruisms” to exist, as well as the possibility for individuals to be altruistic to a group concerning certain goods and nonaltruistic or negatively altruistic concerning others. Finally, it resolves some anomalies of utility-incorporating altruism, such as the implication that altruists will act more selfishly toward other altruists than toward selfish individuals, since other altruists can gain utility by sacrificing their own welfare, or the possibility that individuals who are perfectly altruistic toward one another would have indecisive utility functions, since neither would have self-interested preferences upon which the other could base his or her altruism.41

This implication is consistent with a large number of psychological studies on cooperative collective action. Each study has shown that cooperative action increases liking and tolerance for other individuals, while others have shown that action that hurts others decreases liking toward them.42 As a result, the link between cooperation and liking (the so-called contact hypothesis) has become perhaps the most well-known proposition in the social psychological literature on group

---

41 Rawls 1971, 189.

42 Among the best-known are Allport 1954; Sherif et al. 1961; Aronson et al. 1978. The last may of special interest to collective action theorists because it emphasizes that initial interactions (before altruism develops) in the absence of individual sanctions for noncooperation must involve situations where each individual’s contribution is essential to group success.
behavior. A separate strand of research, more explicitly connected with dissonance theory, has shown that conflictual actions led to increases in hostility toward other individuals. Yet another strand of research, in the social exchange approach, has shown an empirical link between repeated cooperative exchange with other individuals and intrinsic commitment to further interactions with those individuals, as well as a positive relationship between the risk involved and the extent of the commitment.

The inference is also consistent with writings on the politics of identity, which show that while “group consciousness” comes about as a result of collective action, it can persist and affect the levels and boundaries of action far beyond the context in which it was created. Furthermore, these writings note, the strength of the identity created will be related to the importance of the goods at stake in the collective action.

Because cooperation within a group of individuals is often motivated by conflict with other groups, it is also consistent with analysis that posits that conflict can often form the basis for group identity and can strengthen group cohesion. However, conflict can lead to group dissolution rather than cohesion if the existing group identity is not strong enough to ensure cooperative action within the group in response to a threat. In that case, the resulting noncooperative action may lead to a decreasing group identity. A number of studies have pointed to this general effect: during times of trouble, social ties that are already strong tend to get stronger, while those that are weak tend to fall apart.

Proposition: Individuals can reduce doubt by acquiring altruistic preferences toward other members of a group with which they are

43 Katz and Taylor 1988, 3-4. There are actually two interpretations of this hypothesis. The broader interpretation, which associates liking with mere contact, has been subject to a great deal of criticism and at any rate is at odds with Allport’s original formulation. The narrower interpretation is the one used here. For listing of recent work done on this area, see Brown 1986, 613.

44 See Davis and Jones 1960; Glass 1964; Kahn 1966.


47 Huntington and Nelson 1976, chap. 4; Horowitz 1975, 119, 131-32; Hobsbawm 1990, 171; and case studies in Crawford Young, Hobsbawm 1976, chap. 4. For a more philosophical discussion, see Taylor 1992, 118. For further discussions along these lines, see chapters 5 and 6 of this book.

48 Huntington and Nelson point in particular to “intense or sustained conflict” and Horowitz to the brutality of conditions under which cooperation takes place as factors in cementing identity formation. See Huntington and Nelson 1976, 89 and Horowitz 1975, 129-30. See also Smith 1981b.

49 Coser 1956, chapters 2 and 5. These writings were in turn based on the work of Simmel; see “Conflict” in Simmel 1955.

50 Coser 1956, 92-93.

cooperating, if their private return from participation is reliably more positive than the public return.

Cooperation can promote further cooperation not only through preference change but through belief change. In many situations in which a number of individuals are engaging in collective action, it may not be possible for an individual to immediately observe whether others have cooperated, even after the action is completed. However, once the individual has committed to cooperative action, and if an individual’s expected relative utility for a possibly suboptimal action is a positive function of the level of cooperative behavior provided by a group of other individuals (as in a prisoners’ dilemma), he or she can reduce doubt for that action by increasing the expected level of this cooperation. This is simply a special case of an earlier theorem about belief change, since a belief that others will cooperate will increase the expected net benefits from one’s own participation and hence reduce the level of doubt. If such a belief cannot be immediately disconfirmed by information, it can then promote further cooperative behavior.

A final kind of belief change that can occur is to increase one’s perceptions of the contribution of one’s actions to the provision of a public good. When the information assumption does not allow one to directly observe this, it will be doubt-reducing to believe that one’s contribution is highly significant, since it increases the net personal as well as collective benefits associated with the action, an individual can reduce doubt for a possibly suboptimal action that contributes toward a valued shared good by increasing the perceived contribution of his or her action to the expected amount of the shared good.

These inferences establish a key principle, which is that cooperation can be self-perpetuating, even (or especially) in the absence of immediate information about the actions of others or about the ultimate effects of each individual’s own actions. Cooperative action will, through the doubt-reduction mechanism, increase mutual altruism regarding provision of a public good, promote positive beliefs about the cooperativeness of others in a group, and increase perceptions of one’s own contribution to the public good; all can in turn facilitate further cooperation, which can promote further altruism, and so forth.

Collective Action and Shared Ideology
Moreover, the model has some leverage in predicting what sorts of other beliefs and preferences will be shared among members of a group who are each engaged in similar types of actions. Such individuals will often face similar sources of doubt and hence will be able to reduce this doubt through similar sorts of preference and beliefs. For instance, individuals who share a particular profession will have a common need to justify that choice of profession, hence a common tendency to increase their valuation of goods and activities that can be provided only by the profession.

There may of course be a number of preferences and beliefs that can reduce a particular source of doubt and hence a range of possibilities for preference and belief change for different individuals. While this may be so, some of the variation in the ways that individuals within a group reduce their doubt may be narrowed by the existence of
idea entrepreneurs. This type of entrepreneur does not offer selective incentives, but rather supplies ideas about preferences and beliefs to groups of individuals that may be receptive to them. This in turn can lead to commonalities in preference and beliefs within the group.

Although it is becoming acceptable to assert that idea entrepreneurs can shape individual preferences and/or beliefs, there is almost no theoretical writing on how they do so. Entrepreneurs cannot simply by force of their wills impose certain preferences and beliefs on individuals; there must be some psychological equivalent to a demand for the preferences and beliefs they offer. This model suggests that such demand is in general linked to doubt reduction. When a group of individuals face doubt from similar causes and doubt-reducing preferences and beliefs (particularly those that do not increase doubt for past actions) are not obvious, they will be receptive to the first individuals who can suggest and propagate such preferences and/or beliefs. Hence, the entrepreneur’s success in selling his or her ideas is dependent on the fit between the ideas and the doubt problems faced by the potential audience. At the same time, the entrepreneur is not superfluous because he or she can determine from among a set of a potentially doubt-reducing preferences and options those that will actually be adopted.

Entrepreneurs play an especially important role in providing integrated constellations of preferences and beliefs, i.e., ideologies, whether they be political, social, or religious. The most appealing ideologies will be those that provide internally consistent models of the environment sufficiently decisive to make rational decision making possible and that simultaneously lower the doubt associated with actions in a wide range of situations by eliminating perceived tradeoffs in actions. Armed with such ideologies, individuals can reduce doubt without having to resort to the constant and piecemeal adoption of new preferences and beliefs, as well as the possible inconsistencies that can result from them.

Because of the difficulties of developing ideologies that meet these criteria, most individuals are not capable of spontaneously generating them and will rely on those provided by entrepreneurs. If an entrepreneur espousing a particular ideology can communicate with a group of individuals facing doubt problems that can be reduced by this ideology, it is quite likely that these individuals will collectively adopt the ideology offered, even if there are possible alternatives that have not yet been communicated to them.

---

52 See the “appendix” to Olson 1965, 174-76.
54 Hence the manipulative type of entrepreneur will attempt to come up with a set of ideas which will meet her audience’s doubt-minimizing needs, yet, if they are internalized by them, meet her own utility-maximizing needs.
55 This definition of ideology as part prescriptive (preferences), part descriptive (beliefs) accords with that used by a number of prominent writers. See discussion in Boudon 1989, chap. 4.
The origin of entrepreneurs espousing particular ideologies and how they come into contact with certain groups is not something that can be explained by this model or, arguably, by any general model. When this model is applied to the adoption of ideologies (as in chap. 4), ideologies espoused by entrepreneurs need to be exogenized and viewed as part of the environment. In such cases, structural assumptions should be made about the menu of ideologies that are available for adoption. If only one of the available ideologies can reduce doubt for a group of actors, the model will predict that this one will be adopted.

Once an ideology has been adopted by a large proportion of a cooperating group of individuals, there will be strong effects that reduce the chances that these individuals will change their ideologies over time or that other individuals within the group will adopt competing ideologies. Even if two different ideologies solve existing doubt problems equally well, they will at some point have different implications for future behavior. After an individual adopts a particular ideology and begins to apply it to his or her own actions, he or she will eventually commit actions that are contrary to those of a competing ideology, and hence will experience an increase in doubt for those actions if he or she later switches to a competing ideology.

**Proposition:** A group of individuals facing common problems of doubt and multiple ideological solutions will collectively address them by adopting from within these solutions the ideology with the highest cultural salience for all participants.

There are several factors that will promote intragroup unity in ideology, which in turn implies a coherent integration of the ideology into group culture. First, if a group is geographically concentrated or is connected to the same communications sources, it is likely that members of the group will be exposed to ideologies in the same order. Hence, if they are facing similar doubt problems, it is likely that they will end up adopting the same doubt-reducing ideology.

Furthermore, once members of a group have adopted a particular ideology, they will often have an incentive to promote the ideology among other members in order to facilitate future intragroup cooperation. Again, if other members are facing similar doubt problems, they will likely be receptive to this ideology. Furthermore, if a majority of the members of a group adopt a particular ideology and seek to promote it among other members, they will have an incentive to apply sanctions to those members who do not act in accordance with the ideology.

Although sanctioning does not instantly generate internalization of the ideology among those members who initially do not accept it (presumably because they do not share the same doubt problems or have committed past acts whose doubt would be increased by adopting the ideology), earlier discussion in this chapter suggests that sufficient (but not too heavy) sanctioning can lead to gradual preference and/or belief change, particularly in the former case. When sanctions cause individuals to act according to the majority ideology and there is insufficient incentive to act to counter or evade them, one would expect a “sour grapes” effect to promote internalization.
Since sanctioning can only be applied to observable behavior and not internal mental states, enforcement of the accepted group ideology will revolve around a set of rules for actions that are logically implied by culturally internalized preferences and beliefs; these are social norms.\footnote{A similar analysis of the relationship between a group’s defining preferences and beliefs and its social norms can be found in Wildavsky 1991, 316. It is different from the way norms are conceptualized in Elster’s *Cement of Society*, where they are seen as free-standing entities (see discussion in chapter 2).} Enforcement can occur either through centralized authority or metanorms prescribing decentralized sanctioning by group members who observe violation by other members.\footnote{The problematics of decentralized sanctioning are discussed in Milgrom, North, and Weingast 1990.} Sanctioning serves the dual purpose of forcing norm-consistent behaviors among those who have not internalized the ideology and promoting its gradual internalization.

The extent to which attempts at sanctioning will be effective in promoting internalization among members who would not do so otherwise will depend on two factors: (1) the extent to which group members have engaged in norm-violating actions before being subject to sanctioning and (2) the extent to which, under the sanctioning system, effective monitoring can prevent individuals from carrying out violating actions without receiving punishment. For individuals who have already engaged in norm-violating actions, internalization will increase the doubt of these past actions. If individuals can engage in norm-violating actions without being punished, then the internalizing effects of their norm-consistent actions will be offset by the deinternalizing effects of their norm-violating actions.

This means that the individuals most susceptible to internalization through sanctions will be those with short histories of norm-violating actions (hence, no heavy shadow of the past impeding preference and belief change), high vulnerability to sanctions, and little opportunity to evade monitoring. This in turn points out the importance of childhood socialization as the most effective way in which group members can spread and perpetuate social norms that they hold in common. When sanctioning commences at birth, it can be a powerful force for causing the internalization of shared ideologies and hence for the perpetuation of a stable group culture.

While the discussion of socialization is certainly not new to the analysis of culture, previous theories of culture have not provided models to explain when it is effective or how it occurs on the individual level.\footnote{This is particularly true of structural-functionalism, which depends heavily on the existence of socialization processes but provides no clear model of how they occur. See Skidmore 1979, 175-76 and Hechter 1983, 19.} Such models are often criticized for overemphasizing the ability of environments to socialize individuals toward consensually agreed-upon norms;\footnote{Alexander 1987, 122, 131-33.} individuals are seen as “oversocialized,”\footnote{Wrong 1961.} passively
soaking up norms and lacking a free will.\textsuperscript{61} Hopefully, the model presented here provides greater decisiveness than existing approaches to culture-formation do by explaining the origin of cultural preferences and beliefs and specifying the conditions under which internalization takes place.

**Plan for the Following Chapters**

It follows from the definition of doubt that a rational choice model combining the model of preference and belief presented here with the rational optimization assumption can be applied in all environments where the conventional rational choice model can be applied. Furthermore, the coherence rational choice model clearly retains the parsimony of the conventional model. Finally, a number of significant psychological and cultural phenomena, including many outside the range of conventional cognitive dissonance theorizing, can be explained using this model. What remains to be shown is whether the model can address the major empirical indeterminacies and anomalies of conventional rational choice. The remainder of this book is devoted to this task.

The next three chapters apply the model to three substantive issues crucial to the politics of development: economic ideology, ethnic identity, and long-term social change. In each chapter, rational choice and non-rational-choice theories concerning some major political or sociological phenomenon related to these issues is investigated. In each case, it is argued that both conventional rational choice and other theories have failed to provide adequate explanations for the phenomenon, either because of a lack of decisiveness or a lack of accuracy. A new theory is then generated from the assumptions of the model presented here. The theory is shown to explain these phenomena in a way that is deductively grounded yet empirically decisive and accurate. I hope to show via these applications that the model has significant explanatory power over a very wide range of environments and can be considered a useful general model for the psychology of preference and belief formation.

**Proofs of Theorems**

*Proof of Theorem 1*

We can prove this by noting that a rule base in causal form, or any part of it, can be transformed to a formal grammar by transforming the $=$

\textsuperscript{61} See Gellner 1992, 41-54, for a more general examination of the role of socialization in social theory. For rational choice discussions of socialization, see Coleman 1990, chap. 11; 1987; Akerlof 1983; Marini 1992, 38-41; Reder 1979. Each implies that socialization of particular preferences or beliefs will always be successful provided the socializers invest sufficient “resources” into it, then tries to determine the circumstances under which socializers will want to make such an investment. Although this “empty vessels” view of political socialization has largely been rejected within the political socialization field, no comparatively general and decisive theory has been proposed to replace it. See Renshon 1992, 448-52.
(equals) operator in a statement with nonzero probability weight to the → (production) symbol in such a grammar. The resulting grammar can be shown to derive the set of all possible dispositions for an expression that are consistent with the set of rules being translated, and to be decidable. Given this, the problem itself is decidable.

A formal grammar is specified by a set of nonterminals (or “variables,” in yet another sense of the term) representing parts of speech, a set of terminals representing words, a set of production rules mapping strings of nonterminals and/or terminals onto each other, and a special nonterminal for the class of expression being derived.

Part 1: We can begin by considering a foundational syntax that assigns contingencies and wild cards to classes of constants. Let each wild card and attribute symbol (i.e., the main symbol for a subscripted or unsubscripted contingency) translate to a nonterminal, and each constant symbol to a terminal. For wild card symbol $\hat{y}$, create the production $\hat{y} \rightarrow Y$ for each constant $Y$ in its class, and similarly with every attribute symbol. Create the production $\hat{y} \rightarrow y$ for each unsubscripted contingency over the same class as wild card $x$, and $\hat{y} \rightarrow y_i$ for each subscripted contingency.

A formal grammar is context free if the left-hand side of all productions comprises a single nonterminal. It is clear that the productions we have generated so far meet that criteria. To incorporate formation rules for basic arithmetic and logical operators (i.e., constant mappings), productions like $\mathbb{R} \rightarrow \mathbb{R} + \mathbb{R}$ can be included. Many sets of context-free production rules for arithmetic and first-order logic exist. For instance, such production rules are encoded in every computer program that parses such notation.

Note that no separate nonterminal is required for class names, since the wild card can serve as a universal nonterminal for all contingencies and constants in its class. Taking any such wild card as a start symbol, the grammar will derive only those expressions that are within the class of the wild card according to the declarations embodied by the syntactic rules of the foundation. This is because class declarations are transitive in mathematical notation in the same way that productions are transitive in formal grammars. In other words, if $f(Y) \in X$ and $y \in Y$, then $f(y) \in X$ for all $X$ and $Y$ in mathematics. Likewise if $X \rightarrow f(Y)$ and $Y \rightarrow y$, then $X \rightarrow^* y$ in a formal grammar, where $\rightarrow^*$ is the reflexive and transitive closure over $\rightarrow$ and $X \rightarrow^* y$ represents the derivability of $y$ over some sequence of productions, from $X$. Hence as long as each $\rightarrow$ translates to the appropriate $\ni$ for a single production, it follows from mathematical induction that all derived expressions from a wild card are within its class, and that any expression within the class can be derived.

62 In fact, the rules of most compilers translate to an even more restrictive form of specification, the LR(n) grammars, though this is not crucial for our purposes.
From this, proving decidability is simply a matter of pointing out that membership of a statement within a language is inherently decidable over all context-free grammars.63

**Part 2:** Rules in an identity will specify possible mappings more narrowly than those in a foundation. However, they will share the same structure as those in a syntax in being context free. Hence, if we map the rules of an identity onto their own grammar, such that each statement with probability greater than zero becomes a production, the set of statements that are consistent with the identity will be those that can be derived from the grammar.

**Part 3:** The main thing to show for this part is that a solution for the possible dispositions of any contingency (including utility) will exist and can be derived in a finite amount of time through application of a fixed algorithm over rules.

Here, the mere fact that a grammar is context free does not ensure that a solution can be found since the grammar itself may represent an insoluble problem. Under such a grammar, no disposition for a particular contingency may be derivable, not because the rules are contradictory but because no set of productions starting from a contingency can derive an expression of constants.

It is for this reason that causal form is important. Causal form ensures that all contingencies can be placed within a partial order such that the contingencies on the left side of a statement with nonzero probability in a rule base are always posterior to those on the right side. Such a partial order will generate a partition of an individual’s vocabulary such that any contingency can be assigned a number corresponding to the index of its equivalence class in the partition. Given that the number of contingencies defined in a rule base are finite, the number of equivalence classes must be finite as well.

Now, for any expression \( x \), we can assign the number \( \iota(x) \) corresponding to the index of the highest (i.e., causally most posterior) partition of any contingency in the expression. One thing that causal form implies is that any set of productions that transform all the contingencies in expression \( x \) to \( x' \) ensures that \( \iota(x') \leq \iota(x) - 1 \). Given this, there will always be a finite number of steps by which \( x \) can be transformed into an expression \( x_0 \) such that \( \iota(x_0) = 0 \), i.e., \( x_0 \) is an expression containing only constants.

**Proof of Theorem 2**

We assume rule base \( P \) and contingencies \( u \) and \( x_0 \ldots x_N \) in the vocabulary of \( P \), where \( u \) is utility and \( x_0 \) is a contingency deterministically linked to a volition such that the choice set over \( x_0 \) is \( \{x_0 : x_0 \in X_0(P)\} \). For this analysis, it does not matter whether \( x_0 \) is an action or is linked deterministically to an action by the set of

---

63 In fact, membership of a statement within the language of a context-free grammar is not only decidable, but can be determined in polynomial time, i.e., in a time that increases at a rate that is no more than a polynomial function of the length of the expression whose membership is being tested. A well-known example is Earley 1970.
rules. A choice maximizing $x_0$ across all dispositions of fates can be represented by $\dot{x}_0 = \max_{x_0 \in X_0(P)} x_0$.

$S \in \hat{S}(P)$ is a disposition of the vocabulary of $P$. Let $(x_0 \ldots x_{N-1}, u = x_N)$ be a causal path of propositions leading from contingency $x_0$ to $u$ in $S$, with $N \geq 0$. Let $S_N = (s_0 \ldots s_N)$ be the corresponding set of statements. $\dot{x}_0 \ldots \dot{x}_N$ are the corresponding values in this sequence for $\dot{x}_0$.

Let $\hat{x}_{n-1}$ be the lambda abstraction $\lambda x_{n-1} \dot{x}_{n-1}(\oplus_{m \neq n-1} s_m)$, the disposition of $x^n$ as a function of its predecessor on the path, with $\hat{x}_{n-1}$ having a positive derivative and negative second derivative over $x_{n-1}$ for all $N \geq n > 0$. It follows that any perturbation of $s_{n-1}$ that increases $\dot{x}_{n-1} = \max_{x_{n-1} \in X_{n-1}} (S \setminus \{s_{n-1}\}) x_{n-1}$ will increase $\dot{x}_n = \max_{x_n \in X_n} (S \setminus \{s_n\}) x_n$ as well. But then by mathematical induction it follows that any perturbation of $s_1$ that increases $\dot{x}_1 = \max_{x_1 \in X_1} (S \setminus \{s_1\}) x_1$ will increase $\dot{u} = \max_{u \in U(S)} u$ as well. Since $\dot{x}_0 = \max_{X_0(P)} x_0 = \max_{X_0(S)} x_0$, increasing the derivative of $\dot{x}_1$ over $x_0$ will increase $\dot{x}_1 = \max_{x_1 \in X_1} (S \setminus \{s_1\}) x_1$.

Now given that this is true for all dispositions $S \in \hat{S}(P)$, it follows that any such perturbation will increase $\dot{u} = \max_{u \in U(S)} u$ across all dispositions, including the ones in which $\dot{u}$ is suboptimal.

**Proof of Theorem 3**

Let $\dot{u}(S)$ be the level of doubt existing under disposition $S$ of $P$. Then it is quite simple to show that if we increase the subjective probability $p(S)$ under $P$ of $S_1$ by $\Delta b$ and subtract an equal amount from the subjective probability of $S_2$, there will be a change in doubt of $\Delta [\dot{u}(S_1) - \dot{u}(S_2)]$, which will be negative if $\dot{u}(S_1) < \dot{u}(S_2)$.

**Proof Of Theorem 4**

From theorem 2, it follows that the larger the increase in $\dot{x}_{n-1} = \max_{x_{n-1} \in X_{n-1}} (S \setminus \{s_{n-1}\}) x_{n-1}$, the larger the increase in $\dot{u} = \max_{u \in U(S)} u$, since each incremental increase in the former will lead to an incremental increase in the latter.

Given that an individual will always try to minimize the amount of adjustment necessary, it follows that the largest amount of adjustment will be needed when $\max_{S \in \hat{S}(P)} \dot{u}(S)$ is the largest. Decreasing $\dot{u}$ relative to $\max_{u \in U(S)} u$ by a fixed amount $\Delta \dot{u}$ for every $S \in \hat{S}(P)$ will increase $\dot{u}(S)$ for every $S$ by $\Delta \dot{u}$ as well, hence it will increase $\max_{S \in \hat{S}(P)} \dot{u}(S)$.

**Proof of Theorem 5**

The risk factor of a choice is simply a multiplication of the variation in outcomes across dispositions of fate. From the point of view of doubt, this can most easily be accomplished by simply multiplying the individual’s utility function by a constant factor $c > 1$, thus multiplying
\( \tilde{u}(S) \) by \( c \) for all \( S \). The fact that total utility is changed as well does not matter, since totals cancel out when calculating doubt. Then it follows from the logic of theorem 4 that the amount of change in 
\( \hat{x}_{n-1} - \max_{x_{n-1} \in \hat{X}_{n-1}} (S \setminus \{s_{n-1}\})x_{n-1} \) will be increased as well.
Chapter 4
Ideology Formation and Policy Choice in Ex-Colonies

The social science literature is full of works that purport to recommend the correct economic policies for developing countries. However, this wealth of analysis is not matched by theories that attempt to explain the policy choices that Third World states actually make and the reasons for those choices. Positive analysis of economic policy formation in developing countries is a relatively recent academic phenomenon, and theories that attempt to explain variations in development strategies across a wide range of countries are close to nonexistent. Given the obvious centrality of this issue, the absence of systematic explanations for such variations presents a rather gaping hole in the social science literature on development.

This chapter attempts to provide such a theory, drawing upon an endogenous analysis of ideology, which in turn is drawn from the coherence model put forward in chapter 3. More specifically, the analysis is aimed at predicting variations across countries in the level of state economic intervention. This has long been considered the single most politically salient aspect of economic policy, and it is certainly the most widely examined policy variable in comparative studies of economic development.

Despite the prominence of this issue, very few theoretically informed hypotheses have been put forward for predicting levels of state intervention across the full range of developed and developing countries, and they have not inspired a coherent body of literature. This, I argue, is due to the inability of existing theories to account endogenously for perhaps the single most important variable in development policy choice: ideology. Most conventional theories of policy formation ignore ideology. If not, they view it as an attempt by state leaders to rationalize policies serving the interests of the ruling elite or as a set of policy promises for gaining support from self-interested members of society. The possibility that state elites may internalize particular ideologies and use them as a basis for policy formation is denied by the conventional model as well as most modifications of it. One particular literature, the growing body of work on "ideas," does assign a causal role to ideology, but it cannot account for its origins in deductive manner and hence is difficult to use predictively.

Nonetheless, outside the narrow circle of social scientists who design policy-formation models, it is scarcely questioned that ideology

---

1 See for instance Downs 1957, 116; Conover and Feldman 1981, 638.
3 Neo-Marxist theories are among the most prominent that take the former approach. The most notable example of the latter is in Downs 1957, chap. 7.
Choosing an Identity

has been crucial to policy formation throughout history. Furthermore, the extent of state intervention in the economy has been the central policy variable in the great ideological debate between capitalism and socialism. While history books have recorded this century as an epic battle between these competing economic ideologies, both are notably absent as causal factors from conventional theories of state policy making.

I argue that ideology is crucial to explaining development policy choices made by leaders of developing countries, and that it can be accounted for endogenously. The theory presented here accounts for the origins of ideology by tracing it to the experiences of state leaders. It focuses in particular on a particular set of countries: newly-independent former Western colonies. In all such countries, states were initially controlled by leaders who had engaged in earlier pro-independence activities. It shows how, during such activities, any plan of action that involved acceptance of the economic ideology of the colonial power tended to generate greater self-doubt than one in which such an ideology was rejected. Furthermore, it is argued that this rejection arose not so much from a simple desire among pro-independence elites to reject all things associated with the colonial power. Rather, it was linked to the fact that engaging in such activity was largely an all or nothing decision with great risks, which in turn created a set of beliefs that would eliminate any possibility that the decision might be suboptimal. A collective process of identity construction caused elites to internalize preferences and beliefs that magnified the expected diffuse benefits that would arise from independence. Such beliefs in turn could be provided in part by an economic ideology that diametrically opposed the perceived dominant ideology of the colonial power. Historically, I argue, the ideology of socialism provided such a set of beliefs for those engaging in struggles against Western colonialism.

Once independence was achieved, this ideology is posited to have systematically affected the types of policies that the leaders implemented. Since there is no reason to expect similar ideological internalization to occur among leaders of other countries, there will be systematic variations in policy between ex-colonies and other countries, even taking into account structural economic differences between them. However, the number and power of state elites that have engaged in anticolonial struggles will decline over time, and this causes the policies of ex-colonies to eventually become less distinct from those of other countries.

The theory is tested statistically, using pooled cross-sectional time-series data, with regressions for two measures of state intervention, the size of government spending versus the size of the economy and the extent to which prices are distorted from their market levels. Each regression includes not only independent variables representing colonial experiences, but also several independent variables representing alternative explanations for variations in policy that have been taken from existing theories of policy formation.
Theories of Policy Formation

Conventional theories of economic policy formation typically implicitly incorporate a version of conventional rational choice assumptions. The most common set of assumptions takes the state as a unitary entity that rationally optimizes its chances of remaining in power and attributes its beliefs to information, i.e., inference from observations about the society and the international environment surrounding it. However, a number of alternatives have been offered, including such modifications as disaggregating the state and analyzing internal competition among governing elites, placing limitations on the state’s ability to optimize, positing an alternate maximand for the state to pursue, or positing alternate sources of beliefs. Hence, examining these theories allows us to examine the problems that arise in applying conventional rational choice assumptions to policy formation, as well as in modifying these assumptions. Indeed, it will be argued that the problem of policy formation theories is a paradigmatic example of the problem of conventional rational choice: those based on conventional models of the state ignore many of the major factors important to decision-making among elites (particularly those in the developing world), while nonconventional ones have a difficult time generating assertive predictions.

The theories include what is often referred to as the “new political economy,” explicitly rational choice analysis of policy making, but also theories that fall under the older tradition of interest-group-based political economy analysis as well as neo-Marxist and neodependency approaches. Although only the first is commonly identified with rational choice, they are all generally compatible with the assumption that actors are rationally maximizing their material self-interest within a particular set of structural constraints.

For this reason, the following review of the policy formation literature will be organized around different assumptions about the state rather than the usual distinctions made between state-and society-centered or state-, society-, and international system-centered theories. One reason for this is that the main issue we are concerned with, the failure to incorporate ideology into state policy formation, exists among theories in all three categories. Another is the either-or nature of the categorization, despite the fact that all of these theories contain implicit assumptions about the state. Even among theories that have been accused of overemphasizing societal or international interests, the tendency is not to ignore the state altogether but to

---

4 For overviews, see Buchanan, Tollison, and Tullock 1980; Colander 1984; Stigler 1988; Magee, Brock, and Young 1989 and Meier 1991b. Usage of the term seems to date back to Mitchell 1969 and (independently) Ichman and Uphoff 1969, chap. 2.

5 See discussion in the section on structural assumptions and models in chapter 2.


assume structural conditions under which it is in the state's interest to respond quickly to outside pressures.\footnote{8}{For discussion of a similar point, see Almond 1990b and Migdal 1987, 392. A recent classification in Hall 1993 takes this into account, distinguishing between “state-centric” and “state-structural” theories. However, even “state-centric” theories cannot totally ignore the constraining effects on the state of societal factors, as argued in Bardhan 1991, 107.}

I will for similar reasons eschew an attempt to classify existing theories of economic policy formation based on schools of thought.\footnote{9}{Prominent examples include the distinction between functionalist, cultural, public choice, group theory, and state-centric theories in Hall 1986, chap. 1; between normative, values and state/institutional explanations in Steinmo 1988; and between class analytic, pluralist, and public choice (society-centered) and rational actor, bureaucratic politics, states’ interests (state-centered) in Grindle and Thomas 1991, chap. 2; Grindle 1989. With respect to trade policy formation, prominent examples include the distinction between market, cognitive, institutional, and institutional perspectives in Odell 1990; and between rational-actor, organizational and bureaucratic politics explanations for policy-making, distinguishing between rational-comprehensive, incrementalist, political cycle, systems analysis, bureaucratic politics, organizational procedures and public choice approaches. These categories subsume the classic distinction between rational-actor, organizational and bureaucratic politics explanations for policy-making made in Allison 1971.}

Whatever the advantages of such categorizations, what is most important from our point of view is that the dimensions along which schools of thought are distinguished may or may not correspond to differences in assumptions about the choice apparatus of policymakers.\footnote{10}{Bendor and Hammond 1992. See p. 319n.1 of this article for citations to further categorizations of theories of policy-making, based in large part in differences in assumptions about state actors.}

The review begins by examining a conventional model of the state, one that views it as a unitary rational optimizer whose beliefs are logically derived and whose preferences are egoistic and concerned with staying in power. It is argued that such a model underlies most of the prominent general theories regarding variations in levels of state economic intervention, though all are aimed primarily at addressing policy in industrialized rather than developing countries. It then examines modifications of these assumptions based upon disaggregation of the state, bounded rationality, and unconventional about preferences or beliefs. Divided in this way, theories of policy making based on clearly stated assumptions can be classified in a manner that is both exhaustive and assigns each theory to a single category.

Many of these theories focus on several dependent variables, and hence it should be emphasized that no attempt is made here to make a thorough critique of their usefulness or overall explanatory power. Instead, the survey focuses on the types of independent variables that the theories use to account for variations in the level of this intervention,
both in the form of price-distorting protection and/or domestic rent creation and in the form of government spending relative to GNP.

The Conventional Model of the State
Although the conventional model of the state is derived in part from models of electoral competition between political parties in industrialized democracies (the state being equated with the party in power), it has been extended to developing countries by a number of authors who emphasize the need for even nondemocratic states to maintain a support coalition in order to remain in power. Nonetheless, most theorizing that addresses variations in economic policy remains focused on industrialized, primarily Western democracies.

Rent-Seeking and DUP
With regards to levels of state intervention in the economy, and price distortion in particular, the most prominent relevant literature in economics is that on rent-seeking and directly unproductive profit-seeking (DUP). Rent-seeking/DUP theories are concerned with explaining a number of dependent variables such as the amount of resources different sectors will devote to lobbying for government intervention, which policies political parities will support, how voters will vote, and the levels of various types of state intervention in the market. Tariff levels are the most common measure of state intervention. Variations in their overall levels are attributed to such factors as the marginal benefits accruing to each group from expending lobbying resources, national capital/labor ratios, and the degree to which the distribution of factor endowments among voters are asymmetrical around the median.

Of the theories that attempt to explain levels of state economic intervention, one whose independent variables are readily measurable across a wide range of countries is Magee, Brock and Young’s later work on tariffs. They assume only two factors, labor and capital, and that (in developed countries) capital favors an export subsidy on capital-intensive goods while labor favors a tariff on labor-intensive goods. As the capital-labor ratio rises in developed countries, the resources expended on lobbying by capital increase relative to those expended by...
labor, and the amount of export-subsidy increases while the amount of tariffs decreases.\textsuperscript{16} In developing countries, the opposite occurs. This in turn would imply an inverted \textit{U}-shaped relationship between capital-labor ratios and tariff levels.\textsuperscript{17} However, the implications of these hypotheses for variations in levels of overall government intervention are not clear. Not only is the location of the peak of the inverted \textit{U} for tariffs not specified; the inverted \textit{U} is matched by a right-side-up \textit{U} for the relationship between levels of export subsidy and capital/labor ratios. Depending on the exact shape of each curve, the effect of capital/labor ratios on overall levels of protection can take a variety of shapes: monotonically increasing or decreasing, single peaked or double peaked (inverted or uninverted).\textsuperscript{18}

\textbf{NeoDependency and Neo-Marxism}

A alternative and more historical approach to policy formation can be found in neodependency theories of policy formation, which trace state policies that maintain openness to the world economy to a political alliance between the state, the local bourgeoisie, and multinational capitalists.\textsuperscript{19} However, these theories provide little material from which to predict variations in economic development policy among Third World states. Although they do not unequivocably claim that the triple alliance is inevitable, they also do not specify the conditions under which the alliance and its resulting policies can be avoided.\textsuperscript{20}

Traditional Marxist theories do not provide any additional insight into the variations in government intervention since they argue that states uniformly act as agents of the dominant capitalist class. Certain strains of Neo-Marxist literature, inspired primarily by the work of Poulantzas, have challenged this notion and focused on the relative autonomy of the state as a factor in policy formation.\textsuperscript{21}

\begin{itemize}
  \item \textsuperscript{16} Magee, Brock, and Young 1989, 122-26.
  \item \textsuperscript{17} Magee, Brock, and Young 1989, 235. The theory follows a Hechksher-Ohlin-Samuelson model of the economy, hence it has clear links to Rogowski’s work on political cleavages. However, Rogowski posits three factors: land, labor and capital, and focuses on the effects of exogenously determined changes in exposure for each. Owners of relatively scarce factors will favor protection and be aggressive during periods of declining trade while owners of relatively abundant factors will favor free trade be aggressive during periods of expanding trade. He however does not attempt to explain the effects of this on policy. See Rogowski 1989 and Rogowski 1990, chap. 1.
  \item \textsuperscript{18} In his major test of this hypothesis of the “endowment effect,” Magee divides a sample of 58 countries into two groups at the median capital/labor ratio into “developed” and “developing” groups and tests the linear relationship between capital/labor ratios and tariff levels. He finds a statistically significant relationship between the two (with the correct sign) for developed countries and a nonsignificant relationship (again, with the correct sign) within developing countries.
  \item \textsuperscript{19} Cardoso 1973; Cardoso and Faletto 1979; Evans 1979; Gold 1986.
  \item \textsuperscript{20} Packenham 1992, 87-94; Smith 1979.
\end{itemize}
According to such thinking, a relatively autonomous state, because it need not directly obey the dictates of sectoral interest groups, is free to pursue the corporate interests of the capitalist class as a whole. This set of theories begins to blur the line between the conventional and nonconventional models of preferences since they implicitly posit some secondary set of state preferences, class welfare, that is pursued after the primary purpose of staying in power is satisfied. This type of dual utility might be simplified somewhat, however, by adding the assumption that class-welfare-oriented policies enhance the long-term survival prospects of relatively autonomous states, which are free from having to worry about their short-term survival. A plausible interpretation of this argument would be that relatively autonomous states will intervene less in the economy than nonautonomous states.

However, Neo-Marxist writers have not come up with a clear set of hypotheses about the conditions under which states will be autonomous or nonautonomous, nor have they shown how one can go about measuring relative autonomy. Nor is there any clear agreement among them on whether variations in state autonomy should be expected at all. Hence, even this more nuanced Marxist view of the state does not provide many grounds upon which to predict variations in economic policy.

Wealth, Wagner’s Law, and Protectionism
The most prominent hypothesis concerning the size of government is Wagner’s law, which posits that government expenditure as a percentage of GDP will grow as a function of per capita GDP. This hypothesis is generated from a theory that posits that societal demand for government expenditure increases at a faster rate than economic growth, and assumes a state that is responsive to such demands because it is held accountable through democratic processes.

This hypothesis has been tested numerous times in the literature on public expenditure growth, and the statistical results have been mostly (though not always) confirmatory for the industrialized Western countries considered within that literature. However, similar tests that include developing countries have not been forthcoming, and the general accuracy of the hypothesis has not been established.

More recently, Bates and his colleagues have formulated a more complex argument regarding wealth and its relationship to both government spending and protectionism in developed and developing
countries. Their basic premise is that states of wealthy countries and those of poor countries respond to trade instability in different ways. The first set of countries provides compensation in the form of social welfare spending, while the second cannot afford to do so and resorts to protectionism (which in turn distorts prices). Hence, they predict higher levels of protectionism among poorer countries. They also predict that wealthier countries will have higher levels of per capita social welfare expenditure and that per capita social welfare expenditure will be inversely related to levels of protection. Although they do not specify whether they will also have higher levels of social welfare expenditure relative to GDP à la Wagner’s law, their theory does not contradict it.

A similar argument about the relationship between wealth and protectionism can be derived from Krasner’s writing on north-south relations, in which he contends that poorer states are more likely to favor international institutions that promote “authoritative allocation” of goods because their external and internal political weaknesses make them more vulnerable to economic shocks. Although the argument was formulated mainly to apply to international institutions, it also implies that poorer states would be more protectionist. However, unlike either Wagner’s law or the article by Bates et al., it might also imply that poorer states would also favor larger governments expenditure as a proportion of GDP, though whether they could realize it or not is another question.

**Small States, Openness and Corporatism**

Although neoclassical trade theory indicates that liberal trade policies are optimal for all countries, the economic and political costs of protectionist policies may be especially high for small economies since they may not be able to attain the economies of scale necessary for

---


28 In probit regression involving 32 countries, Bates, Brock, and Tiefenthaler 1991 found a significant positive relationship between the log of per capita transfer payments or per capita revenue and World Bank classification for protectionism (augmented by their own classifications for open European economies). They did not perform a direct test of the relationship between per capita GDP and protectionism, although an earlier test by Conybeare 1983 found a significant positive relationship of per capita GDP with 1971 tariff levels, though not with 1902 tariff levels. Another regression by Bates, Brock, and Tiefenthaler 1991 tended to show a positive relationship between the log of terms-of-trade instability and the classified level of protectionism, although one problem with this test is that it does not incorporate any feedback causal link between government policy and instability. Since successful protectionist policies will not only reduce the overall amount of trade but also change the factor breakdown of exports and imports toward those whose prices are more stable, instability would appear to be in part a consequence of existing levels of protectionism. In that case, their theory would not imply that governments would set tariff levels according to existing levels of instability but rather some estimate of what instability would be absent existing levels of protection.


30 Krasner 1985, chap. 2. For a criticism of this position, see Srinivasan 1991, 141-42.
efficient production in all sectors. Hence, they may be more likely to counter the political costs of openness with compensatory social welfare policies rather than protectionism. The most highly developed form of such analysis can be seen in the literature on the small economies in Western Europe, in which Katzenstein argues that states in such countries simultaneously implement open trade policies and corporatist domestic economic policies. According to Katzenstein, these domestic policies include income supports, employment subsidies, unemployment insurance, and retraining as well as generous spending on education and health. All this implies that governments in small states impose heavy tax burdens and have control over a large portion of national income.

Certain interpretations of this theory have taken openness (levels of exports and imports as a percentage of GDP) as an independent variable rather than size of the economy. To the extent that open trade policies promote openness, these interpretations argue that one type of intervention (protection) can be replaced by another (big government).

System Age and Interest Group Gridlock
Olson’s theory of rise and decline provides a political economy version of the old adage about the shrinking economic dynamism of states as they grow older. According to this theory, stable societies with unchanged boundaries slowly develop “distributional coalitions,” organized interest groups that seek government intervention in their favor. In turn, the growth of distributive coalitions “increases complexity of regulation [and] the role of government,” which in turn inhibits economic growth. This theory implies, then, that government intervention should increase over time as long as states do not undergo major social or political upheavals. A similar time-dependent model of governmental growth, though with a different explanatory basis has been formulated by Lindbeck.

---

31 Katzenstein 1983, 1985 and Katzenstein 1984. The first two are theoretical works, while the third offers in-depth case studies. See also Lindbeck, e.g., Lindbeck 1974 and Lindbeck 1976. With regards to trade policy, Conybeare concluded that size was not a good predictor of recent tariff levels among the developed countries. He finds greater confirmation for developing countries, although his sample does not include African countries. See Conybeare 1983, 459-60.  
32 Small States, 47-56.  
33 Cameron found a significant positive relationship between openness and growth in the size of government, while in a more recent test Hicks and Swank found a significant positive relationship between openness and size of welfare spending as a proportion of GDP. See Cameron 1978 and Hicks and Swank 1992. See Conybeare 1982 for a related test, although the assumptions about the state are different.  
34 Olson 1982.  
35 Olson 1982, 24.  
36 The only large-scale test of this theory, by Pryor, used five measures of economic growth as dependent variables rather than levels of government intervention. See Pryor 1984.  
37 Lindbeck 1983.
Arguments of this sort can be found in relationship to social welfare spending in particular. Wilensky has argued that commitments to social welfare programs are very difficult to withdraw once in place, and hence the amount of social welfare spending ought to go up with the time that has elapsed since such programs were put in place. This argument implies, as does Olson, that government spending as a percentage of GDP will tend to increase as a function of the age of a regime.

Other Hypotheses
The survey has covered the major bodies of theoretical work on policy formation, but it has not discussed all the possible hypotheses that have been put forward to explain variations in state economic intervention. In the literature on why government grows, the list of suggested variables that might influence policy is huge, though not all of them are derived from clearly stated theories. However, many of the variables that have been suggested cannot easily be applied to a wide cross section of both developed and developing countries. Some only apply to functioning democracies such as those relating to “electoral business cycles” or the difference between parliamentary and presidential systems or between proportional and majoritarian systems. Others, such as extent of fiscal centralization or extent of reliance on indirect taxation, are simply not available for many developing countries. Others, such as union strength, presuppose some minimal level of industrialization and democratization. In sum, the range of readily measurable variables that can apply to developing countries is considerably smaller than that which can be applied to developed countries.

Variations on the Conventional Model of the State
Discontent with the adequacy of the explanations offered by the conventional model of the state has led various authors to offer alternative views. One prominent approach seeks to disaggregate the unitary view of the state and investigate the actions of the lower-level actors that comprise it. The other generally retains the unitary state but abandons the view that it is a rational optimizer, instead positing either nonrationality or some sort of bounds on rationality. Other approaches retain the assumption of rational optimization but alter the conventional assumptions about preferences and/or beliefs.

38 Wilensky 1975.
39 For a linear regression analysis that includes just about every conceivable variable that might influence government social welfare spending, see Hicks and Swank 1992.
40 Most notably Tufte 1978.
42 Hicks and Swank 1992.
43 e.g., Gertz 1981; Conybeare 1982.
44 e.g Cameron 1978 and Conybeare 1982.
45 e.g., Alvarez, Garrett, and Lange 1991.
Disaggregated State Theories
As a number of authors have noted, states in the real world, however they are demarcated, are hardly unified but are comprised of individuals with autonomous preferences and beliefs. While viewing the state as a unitary actor may be a necessary shorthand for reducing the complexity of analysis, such analysis may lead to anomalous predictions unless there is some reason why the actions of individuals who comprise the state would converge around a similar set of goals with regard to the domain of policy being investigated.

Among theories of foreign policy formation, there have been a number of well-known attempts to apply a perspective that emphasizes the infighting and clashes of interest that take place between groups individuals situated in off different branches of a government. Furthermore, there are a number of theories that apply the notion of a divided or factionalized state to economic policy-making. However, these theories have not produced any clear hypotheses about the factors that cause variations in levels of state economic intervention across different countries.

There are two ways in which policy variations across countries can be explained within a disaggregated politics model. The first is through intragovernmental variations in the beliefs of policymakers and/or the maximands that they are pursuing. But mere disaggregation does not provide any insight into the content of such preferences and beliefs; this requires an additional methodology (which will be discussed later). The other way in which policy variation can be explained is through variations in the organizational structure of the state, and it is here that disaggregated state models may be able to provide some insight. However, organizational explanations have a somewhat incomplete quality about them in that the organization itself is endogenous to political action. Given rationality on the part of political actors, any policy-significant variation in the organizational structure of different states ought to reflect differences in the preferences, beliefs, or relative capabilities of the political actors who originally created them. Although the impact of an organizational structure may persist longer than the political environment that originally fostered it, it is true nonetheless that organizational explanations should in principle be traceable to broader political or socioeconomic explanations.

Bounded Rationality Theories
Another way in which the conventional model can be perturbed is by positing some bounds on rational optimization. This can be done either by retaining the state as a unitary actor or by also disaggregating the state and examining bounds on the reasoning of individual policymakers. In practice, the distinction between these

---

48 Most notably Niskanen 1971.
can be vague. The usual tendency in theories of policy formation is to fudge this distinction by equating the state with a small group of policymakers with similar preferences, beliefs, and decision-making mechanisms with regard to the issue being examined, which in turn cause them to evaluate policy options in a similar manner, choosing policies as if they were a unitary entity.

There have been very few attempts to systematically apply bounded-rationality actor models to policy making, the most prevalent being various attempts to apply the prospect theory of Kahneman and Tversky to crisis decision making. Such applications focus on explaining single cases rather than cross-sectional analysis, and no attempt that I know of has been made to apply bounded-rationality models such as prospect theory or satisficing cross-sectionally to explain variations in economic policy formation. In order to explain variations in policies, such an application should be able to explain why different sets of policymakers should have different bounds on their rational decision making, and how these differences systematically effect levels of state intervention in the economy. Such a theory seems quite difficult to formulate without somehow taking into account the variations in cognitive abilities of different decision makers along some dimension, which in turn seems quite difficult to determine.

The main way to apply a nonrational actor model to policy formation is by positing that state leaders, or the state as a whole, create policy by following established operating procedures. Variations in policy among different states can then be attributed to differences in their operating procedures. However, this leads immediately to the question of what determines differences in operating procedures. The only recourse in this model is to refer to higher-level operating procedures, which in turn must themselves be explained. At some point, goal-oriented behavior must be brought back in to explain the ultimate source of variations in policy.

**Modifying State Preferences: Alternative Maximands**

Rather than changing their level of analysis or conception of rationality, most theories of policy-formation that do not employ the conventional model alter its assumptions about state preferences and/or beliefs. I first discuss theories that employ nonconventional assumptions about state preferences, arguing that they do not provide a way of accounting for the presence or absence of preferences and/or how variations in

---


50 This of course, does not eliminate collective action problems in actually implementing these policies.

51 For example, see the special issue of *Political Psychology* 13, 2 (June, 1992). For a discussion of prospect theory, see the section on nonexpected utility theories in chapter 2.

52 e.g., “Model II” discussed in Allison 1971, chap. 3.

53 Established operating procedures are a kind of social norm, and generate the same theoretical problems. See discussion on social norms in the section on nonrationality theories in chapter 2.
There are three major groups of such theories. The first stems largely from the theoretical analysis of states in newly industrializing countries and posits that such states are primarily interested in promoting national economic development. On the other hand, another set of theories on policy formation in developing countries focuses on “predatory” states that maximize their own wealth without regard to the economic welfare of their societies and even at the cost of their own long-term survival. Another set of theories focuses on states that maximize neither national nor state economic welfare but instead maximize “nationalism,” which generally refers to some sort of indigenous self-reliance.

**Developmentalist States**

One major facet of the revival of the “neostatist” approach to policy analysis\[54\] has been an increasing focus on the role of the state in promoting economic development or security in both developed\[55\] and newly industrializing countries.\[56\] Implicit in such analysis is the assumption that such states have an overriding concern for furthering some measure of national economic welfare.

In this literature, the main variable that is used to account for variations in economic policy is the autonomy of the state. Ceteris paribus, the more autonomous the state is, the more developmentally-oriented its policies will be. Nonautonomous states, even with development goals, will be so “penetrated” by narrow societal interests that they will not be able to carry them out.\[57\] Unlike Neo-Marxist writing, this literature emphasizes the preeminence of the autonomous state and hence its developmentalist urges rather than its need to serve class interests in order to maintain power. Furthermore, there is a clear consensus among neostatists that there are variations in the autonomy enjoyed by different states and that these variations have a significant impact on policy.

However, like the neo-Marxists, the neostatists have not provided any clear independent way of measuring state autonomy or theories to account for it endogenously. Furthermore, the usual hypothesis found in this literature is that greater autonomy leads to more effective state economic intervention rather than less, and hence the implications of this literature for the variables being examined here is somewhat unclear.

---

54 For overviews of the approach, see Krasner 1984; Skocpol 1985; and Levi 1988a. For a critique of the approach, see Almond 1990b.

55 For the U.S., see Krasner 1978. For Japan, see such “revisionist” works as Johnson 1982; Prestowitz 1988; Von Wölfers 1989; Anchordoguy 1989.

56 e.g., Deyo 1987; Amsden 1989; Gereffi and Wyman 1990; Haggard 1990; Wade 1990.

Predatory States
The flip side of the developmentalist assumptions can be found in theories of the “leviathan” or “predatory” state, which assume that the state is seeking to maximize its own income rather than the income of the society it rules.\footnote{58} Most of these theories focus on pointing out the existence of incentives for such a state to intervene in the economy rather than to predict the amount of intervention that will occur. However since they must also accommodate the fact that states do not simply expropriate the entire available income of their country, many posit that there is some optimal level of confiscation beyond which revenues will diminish because of capital flight, withdrawal from the formal economy or tax avoidance.\footnote{59} However, efforts to specify the variables that determine the optimal level of confiscation have been sketchy. Levi points to the state’s relative bargaining power over control of resources (political, economic, and coercive); the transaction costs of bargaining, measurement, monitoring, agency (monitoring those who monitor) and enforcement; and the state’s discount rate. However, she does not attempt to turn this list of factors into an assertive model of intervention.\footnote{60} Findlay formalizes such effects but only at a very abstract level, positing that the state’s incentives to tax and conscript labor depend on a labor supply curve whose marginal supply is negative beyond some optimal rate of taxation.\footnote{61} Without some effort to specify endogenously the determinants of this labor-supply curve or to show how its parameters can be measured, this type of theory remains difficult to apply empirically to predict variations in state economic intervention.\footnote{62}

Nationalist States
Finally, a smaller body of literature attempts to deal with evidence that states engage in behavior that seems neither to be directed at developmentalist objectives nor to follow from the pursuit of predatory self-interest.\footnote{63} These theories posit that states generate economic policy to maximize some form of “nationalism.” Seers defines nationalism as


\footnotetext{59}{Lal 1987, 281.}

\footnotetext{60}{Levi 1988b, chap. 2.}

\footnotetext{61}{Findlay 1991, 25-26.}

\footnotetext{62}{Recently, Evans has proposed an explanation that attributes the differences between autonomous developmentalist states and autonomous predatory state to their relative degree of “embeddedness” in society. Embedded states, he argues, will respond to the overall concerns of society rather than to a small group of private actors. However, he does not clearly specify the underlying motives behind a state’s activities, nor does he provide a clear way of determining the level of a state’s embeddedness or autonomy. See Evans 1992 and Evans 1995.}

\footnotetext{63}{Johnson 1965 and “The Ideology of Economic Policy in the New States,” in Johnson 1967, 124-41; Seers 1983.}
the urge to promote the presumed interests of a group with cultural coherence, probably showing at least a degree of linguistic and ethnic homogeneity, and usually inhabiting a political unit, or nation-state (though sometimes applied to a group of the same kind submerged within one or more nation-states).  

Johnson defines it as attaching utility or value to having certain jobs held or certain property owned by members of the national group rather than by nonmembers of the national group.

Such theories, unlike most rational choice theories, can provide plausible explanations for phenomena such as redistributive policies favoring members of indigenous ethnic groups over outsiders and public investments aimed at promoting economic self-sufficiency regardless of comparative advantage. However, they do not provide the basis for predictive theory. In order to do so, they would have to specify more clearly exactly where nationalism comes from, and how its nature and level varies from country to country and over time.

One common characteristic of the theories in the above three categories is the fact that they do not claim that their assumptions about preferences are universal. Not all states will be developmentalist, predatory, or nationalist, only a certain subset. However, they also do not provide any means for determining the nature of a particular state’s preferences. Meier expresses the problem this way:

To say that the state is a rational actor of its own utility, without further specification, is vacuous, but the possible variety of utility functions is extensive, and many cannot be predicted but only identified ex-post.

However, ex-post identification alone leaves the rational choice approach tautological and makes it unusable as a means by which to make predictions about policy formation. One way around this might be to search, as Krasner does, for long-term, stable patterns of state behavior and to determine from them the nature of state preferences. If it is assumed that such preferences will remain stable in the future, they can be used as a basis for prediction. However, in situations of frequent government change, it may be difficult to extract stable patterns of behavior or credibly assume that preferences will remain stable in the future.

---

64 Seers 1983, 9.
65 Johnson 1965, 176.
66 These topics are addressed in a different context in the next chapter.
67 Meier 1991a, 302-3.
68 He finds preferences lie “either with general material objectives or with ambitious ideological goals related to beliefs about how societies should be ordered.” With regards to American policy toward raw materials, he posits that the state was concerned, in increasing order of importance, with “increasing competitive economic behavior, ensuring security of supply and furthering broad foreign policy objectives.” See Krasner 1978, 11, 14.
One way to deal with the plethora of possible state motivations is to posit that many of the possible objectives of states can be complementary, at least in the long run. Nationalism may be believed to promote economic development, which may in turn promote tenure in office, which may in turn promote the wealth of state leaders. However, in order to integrate different assumptions about possible state preferences a theory must assume that state leaders in fact believe in such a virtuous cycle of causation. Furthermore since none of the above categories provides clear predictions for variations in preferences, an integrated set of assumptions will not provide such predictions either.

Modifying State Beliefs
The most prominent critique of the conventional model of the state attacks conventional, information-based assumptions of belief formation and emphasizes the impact of ideology on state decision making. As was noted in chapter 1, conventional assumptions about belief posit either that all actors possess all relevant facts about the environment around them or that beliefs are a product of features of the environment that are directly observable to each actor and the logical/statistical inferences that can be made from them (these correspond to complete and incomplete information models). These assumptions are hence based entirely on “objective” forms of belief formation and leave no room for ideology or other factors that might cause similarly situated actors to have different sorts of beliefs. Such a model hardly accords with the usual public view that ideology plays a major part in policy formation. Moreover, many prominent theorists of policy formation have accepted in principle that objectively formed beliefs do not provide a sufficient basis for rational policy making, even though subjective factors remain outside the scope of most positive theory building. Several quotes can provide some hint of the extent to which ideology in particular is accepted and the roles that it is seen to play:

Leaders are often faced with competing causal models for the effects of their policies, and they have no clear way to determine which is correct:

In fact, policy elites may have little direct information on the interests of particular groups in society or on the limits of tolerance for policy actions that do not directly benefit, or even harm, these interests.69

There is considerable ambiguity about economic reality, and ambiguity permits different interpretations. Different understandings or models of a situation can lead to different ends calculations of the costs and benefits of action, its opportunities and disadvantages, and hence of behavior.70

In such cases, beliefs derived from ideology may form the cognitive basis for rational action. As defined in chapter 3, ideology consists of a coherent and explicitly stated set of beliefs and/or preferences

---

concerning a particular domain of human interaction. Rather than being contrary to rationality, such an ideology may be a necessary precondition for rational decision-making to occur. In such a case, understanding actors’ ideologies will be a necessary part of predicting the choices they will make.

The range of options during a time of crisis is not entirely given by the setting. If state elites periodically achieve autonomy from societal influences, the economic ideologies available to them might loom large as an explanation of state action.\textsuperscript{71}

It is difficult for particular groups to calculate where their interests lie. Ideological struggles therefore can outweigh competition among organized interests as a determinant of policy change.\textsuperscript{72}

No theory of institutions would be complete if it excluded ideology. The reason is straightforward... We begin by recognizing that ideology is ubiquitous: individuals have a need to rationalize the world around them. In the sense, ideology, as a device by which to deal with the multiple problems of everyday living that confront one, economizes on the amount of information that people must have. But it is more than that—it also involves a judgement about the fairness or legitimacy of the contractual or institutional arrangements within which individuals live and act.\textsuperscript{73}

Ideology may not only coordinate expectations and delineate legitimate modes of interaction between state institutions and societal actors, but it may also serve as a basic source of identity, and its preservation may be a consumatory function of the state.\textsuperscript{74}

Evidence for the importance of ideology in policy making lies in the fact that real-world actors often attempt to use ideology to shape the minds of policymakers.

If ideology is not important, then economists must explain the enormous amount of resources that political units and other principals in political economic activity devote to attempting to convince participants of the justice or injustice of contractual arrangements.\textsuperscript{75}

Leaving out a role for ideas makes it unclear why economists should attempt to formulate normative analysis in the first place since it cannot have any effect on policy.\textsuperscript{76}

\textsuperscript{71} Haggard 1990, 46.
\textsuperscript{72} Bates and Krueger 1993, 456.
\textsuperscript{73} North 1984, 34.
\textsuperscript{74} Krasner 1984, 228.
\textsuperscript{75} North 1984, 39.
\textsuperscript{76} Meier 1991a, 300-301.
Choosing an Identity

The Ideas Approach

Despite this general acceptance of the significance of ideology, most existing studies of policy formation do not take ideology into account. The major exception to this is the large literature on ideas and their effects on the decisions of state policymakers.

Analysis in this approach is directed at a small group of policymakers, an epistemic community, that shared a common set of causal beliefs. It usually investigates a particular historical case of policy making, providing historical evidence such beliefs were not derived from available information about the policy environment and that they played a role in the policy decisions that were ultimately made. In comparative case studies, an attempt is made to show that elites in different countries held opposing beliefs, and that this led to different policy decisions. The approach is often presented as an alternative to rational choice, but the form of rational choice being criticized is generally the conventional one, and it is clear that the ideas approach can be compatible with the rational optimization assumption.

The terms ideas and ideology are generally used interchangeably in the ideas literature, though the focus in this literature is generally limited to ideologically-derived beliefs. Likewise, given the close links between ideology and culture, the “ideas” approach, or at least a large subset of it, is sometimes referred to as the “cultural” approach.

Taken as political history, these works provide ample and convincing evidence of the impact of ideological beliefs in policy formation in particular cases. However, questions have been raised about whether the approach is falsifiable or whether invocation of ideology is merely a dubious way of arbitrarily accounting for historical phenomena that cannot otherwise be explained. The latter accusation is overstated given the amount of effort that most “ideas” authors make to historically document their claims about the ideological beliefs of policymakers. However, it is true that the ideas approach as a whole cannot be falsified because it does not contain any clear set of assumptions about the causes of ideological beliefs nor a systematic methodology that restricts the kinds of inductive interpretations that can be made about such beliefs from historical data. Because of this,

77 For a dissenting note, see Rogowski 1988, 314-15.
81 As discussed at the end of chapter 3.
82 Hall 1986, 8-10 and Rohrlich 1987, 67-72.
the approach does not account for the origins of ideology,\textsuperscript{84} and hence cannot easily provide a basis for predictive theory.

Certainly, there is considerable discussion in the ideas literature about the types of factors that might influence the choice of ideology such as the influence of intellectuals, the structure of the state, and, most notably, collective learning from past experience. However, the role of these factors has not been specified in a general and deductive manner that can be used as a basis for making predictions about which ideological beliefs will be adopted by policymakers in different cases.

In particular, for learning effects to be part of a deductive theory the theorist needs to know what the unintended outcomes of policy choices will be. This implies that the theorist in a sense needs to know more about policy-relevant matters than the policymakers themselves do. Even if this is the case, for the social learning approach to be more than a set of ad hoc propositions, this superior knowledge must flow from a general set of assumptions. Furthermore, even when the theorist can predict the unexpected consequences of policy choices, she also needs how state elites will react to them. This is no simple matter, especially since there is always the possibility that elites may learn the wrong lesson.\textsuperscript{85}

While it is quite plausible that learning effects will result from significant historical events, there will be variations between countries and individuals as to the extent to which this “lesson” is absorbed. For instance, the reality of relative economic decline in most socialist countries is indisputable, but it cannot be equated with a clear “falsification” of socialist ideology as a whole, one sufficient to create a broad belief consensus among elites on the effects of economic intervention. First of all, this decline can be ascribed to exogenous forces (particularly the depredations of hostile capitalist powers) rather than the failure of economic policies based upon socialism. Second, even when policies are said to have failed, it can be claimed that these policies represented the “wrong” type of socialism and that the “right” type will lead to the desired results.

Although various other mechanisms have been proposed by analysts of ideology to describe why internalized ideologies can come into being,\textsuperscript{86} none has been stated in a manner that can be used

\textsuperscript{84} Goldstein and Keohane state: “We do not seek to explain the sources of . . . ideas; we focus on their effects.” See Goldstein and Keohane 1993, intro. ch., 7.
\textsuperscript{85} See [Hall.,QQPSZ], 293, note 20.
\textsuperscript{86} Most notably “strain” theories, which associate ideologies with rapid socioeconomic change. See Sutton 1956, especially chapter 15; [Shils.,QIZ]. Geertz isolated four ways in which ideologies can address strain: catharsis, the morale-boosting, solidarity-creating, and explanation of underlying discontents. See Geertz 1964, 54, chap. 8. Using a slightly different approach, Lane proposed a paradigm in which ideological change is attributed to exogenous changes in existential base (socioeconomic conditions), cultural premises, personal qualities, and social conflicts, each of which is to a limited extent autonomous from the others. See Lane 1962, chap. 25. In a completely different vein, Boudon focuses on four different cognitive effects, which he labels positional, dispositional, communication and e (epistemological). See Boudon 1989, especially chapter 5.
Choosing an Identity

make clear predictions about the ideologies that will be shared within particular political groups under particular circumstances.

One way to systematize the measurement of ideology is simply to derive it from the stated ideological affiliations of political parties. The most obvious drawback of this methodology is that it can only be applied to countries with well-established political parties and clear ideological divisions. To date, such a methodology has been applied cross-sectionally only to (primarily Western European) industrialized democracies, and the resulting measures have been used as independent variables in multivariate linear regression models of government spending levels. As is usual with such models, statistical estimates of the impact of party ideology and strength on spending tend to vary widely.87

Even if a more general measurement methodology is found, such an approach will be difficult to apply predictively (except in the very short term) because it cannot account for the origins of ideological beliefs or how they change over time. Any attempt to construct a predictive hypotheses about ideological beliefs and how they affect policy making will require a model that can account for ideology endogenously. The following sections attempt to provide such a model, and use it to predict variations in levels of state economic intervention, both between countries and over time.

An Identity Formation Theory of Oppositional Ideology Formation and Policy

The theory presented here is aimed at accounting for variations in the level of state market intervention between different countries and within countries over time. It will do so by attributing them to variations in the internalized economic ideologies of elite groups in control of each state. These ideologies in turn are traced back to the varying political experiences of these elites, and the effects of these experiences on the types of beliefs that were incorporated into their constructed identities during those experiences.

The theory focuses in particular on former colonies, positing that participation by indigenous elites in pro-independence activities created tendencies among such elites to internalize economic ideologies that maximized the perceived benefits of achieving power. Such ideologies had the common characteristic of being diametrically opposed to the ideology that was associated with the colonial power’s policies, and I refer to them as oppositional ideologies. These in turn affected the policies implemented by newly independent states dominated by such elites in the postindependence era.

Oppositional ideologies are posited to create a systematic variation in the types of policies implemented in ex-colonies and other countries.

---

87 Among those who have found a relationship between left party power and levels of spending include Castles 1978; Cameron 1982; Hicks and Swank 1992, and, with certain qualifications, Blais, Blake, and Dion 1993. Wilensky on the other hand has found left party power to be insignificant and Catholic party power to be a better predictor of levels of welfare spending. See Wilensky 1981.
In particular, the theory will focus on the extent to which different states intervene in their economies, positing that states in ex-Western colonies will be more prone to intervene in their economies than those in other countries, even controlling for variations in economic structural conditions. However, the effect of colonial experiences on the ideology of state elites and in turn on state policy is posited to decline over time, as the elites who engaged in anti-colonial struggles are gradually replaced by those who had no such experience and hence no similar reason to internalize the same ideology. This effect in turn will exist when controlled for worldwide trends in economic policy. The predictions provided in this section are tested statistically against two measures of state intervention in the economy, one for the effect of state policies on prices and the other for the size of state spending as a proportion of total GDP.

As is clear from this description, no attempt is made to provide an exhaustive list of all variables that might influence economic ideology or policy (this would undoubtedly be an extremely long list), rather the intention is to isolate one particular set of variables and argue that they have had a significant influence. Nonetheless, because it accounts endogenously for internalized ideologies by tracing them back to the experiences of elites, the theory provides insights into the origins of ideology and how it might change over time. Furthermore, it can be applied more generally as a way to predict the types of ideologies that political actors taking part in antiregime political struggles will find appealing, as well as the types of policy choices they will make if these struggles are successful. Finally, by systematically incorporating ideology into explanations of policy, it can account for variations in levels of state intervention that have been left unexplained by existing theories of policy formation.

Conflict over Sovereignty, Nationalism, and Economic Ideology

In chapter 3, implications of the identity formation model for the causal links between group cooperation, conflict, and ideology were described briefly. These links can be drawn more concretely here.

Among the propositions inferred from the model were that when members of a group are engaged in collective action, and when short-term costs exceed short-term benefits, participating members will suffer from common sources of self-doubt regarding their decisions. Moreover, despite a wide range of conceivable preferences and beliefs members can adopt in order to reduce the level of doubt, internalization of different members will converge to become consistent with a single focused and internally consistent set of preferences and beliefs among those on the publicly available ideational menu as long as there is one such set, ideological by definition, that is unambiguously most prominent amongst those alternatives that addresses shared identity problems, i.e., that reduce self-doubt for the decision to participate in collective action, are consistent with but not determined by available information, and that do not rule against other past committed actions. In order to do so, such an ideology focuses on goods resulting from collective action that are reliably and widely distributed as well as being desired by all group members.
It follows from the coherence model, along with uncontroversial assumptions about structure and initial preferences and beliefs, that members of groups in prolonged conflict with one another over state control will have a tendency to internalize opposing ideologies concerning the policies that will maximize mutually desired objectives. It should be made clear that I am not claiming merely that opposing ideologies can be the cause of political conflict but that political conflict will cause conflicting groups to internalize certain types of opposing ideologies as well.

Given any situation in which there is a disharmony of interests between two groups, there will be a choice for members of each group between conflict and conciliation to resolve this disharmony. It is reasonable to assume that there will be unpredictable and sizable costs and benefits to either choice, particularly when the conflict takes place outside of any stable and predictable institutional structure. Hence violent conflict will typically tend to generate more such unpredictability than conflict taking place within democratic institutional boundaries.

When the sign of relative payoffs is unpredictable, any available strategy will degrade coherence by generating doubt. In this situation, individuals can reduce doubt by internalizing a set of beliefs that reliably increases the relative payoffs generated by the choice that is already seen to have the highest expected utility as long as this set of beliefs is internally consistent, does not conflict directly with phenomenological experience, and does not contradict existing beliefs that have been committed to through past actions.

When all members of a group face a similar source of doubt but there exist multiple, mutually contradictory sets of new beliefs that could serve to reduce this doubt, their internalization will converge to be mutually consistent with one of the choices on the common-knowledge ideational repertoire, i.e., alternative explicitly stated and internally consistent sets of beliefs that have been subject to acts of communication that are publicly available to all. Such sets of beliefs constitute ideologies by definition. If multiple ideologies exist that are consistent with a set of doubt-reducing beliefs, then they will focus on one if it is unambiguously most prominent within the ideational repertoire of group members than all others that also meet the previous two criteria, where prominence is measured by the frequency with which the messages about the ideology are received by members of the group. Moreover, the overall implication of the model is that an internalized ideology will also affect future choices, even when those choices are not directly related to the one that prompted this internalization. Since members of a group participating in a conflict will have common tendencies to internalize certain types of ideological belief, conflict can
Ideology Formation and Policy Choice in Ex-Colonies

be seen as unifier of group ideology, a constructor of common beliefs among individual group members.  

For a situation in which members of an indigenous group are engaged in collective action in order to wrest state control of their territory of residence from members of a nonindigenous group, emergent ideologies will therefore prescribe beliefs and preferences that address identity problems linked specifically to participation in such action. Such beliefs and preferences clearly correspond with the types of ideologies that are usually referred to as nationalist. However, they provide a microlevel grounding for the specific content of nationalist ideologies as well as the consequences of nationalism. Furthermore, they account for certain peculiarities in the nature of nationalism that are rarely subject to explanation, including aspects that clearly contradict conventional rational choice assumptions.

Given this, the relative costs and benefits to members of the opposition group for participation are linked to their beliefs regarding their importance to the groups' chances of success, the public benefits arising from group success, and the private benefits arising from participation as well as their preferences. Hence, the decision to participate in group conflict implies that a rational member believes that the group has a chance to succeed and that he or she can make some positive (though not necessarily large) contribution to this success.

As was conceded in chapter 3, it is usually necessary to make modal assumptions about the initial preferences and beliefs individuals bring to a situation being analyzed, using criteria of reasonability and consistency with past actions, and then focus on predicting the direction of change prompted by the need to eliminate self-doubt. It is reasonable to assume that the modal member of the opposition group possesses an initial utility function such that participation in conflict in and of itself is costly, hence that positive net benefits depend on group success in the conflict. Such an assumption can hold whether or not a member possesses some level of positive altruism toward other members of the group. Structurally, we can assume that she does not possess sufficient information to infer with certainty the overall probability of group success or the relative position of herself or significant others in any new state regime if the group is successful.

Given these assumptions, it follows that a member will experience some level of self-doubt over the decision to participate, and will reduce self-doubt by internalizing beliefs, within the phenomenological and commitment constraints, that maximize the perceived public benefits to group members of state control as long as they can be combined with beliefs that, within similar constraints, maximize the member’s perceived contribution to group success and her private returns from

88 These generalizations are clearly related to some of the propositions explored in Coser’s classic Coser 1956, especially chapter 5. However, unlike Coser’s discussion, which emphasizes efforts on the part of group leaders to enforce intragroup ideological conformity under conditions of outside conflict, it shows why conformist ideologies might in fact be internalized by group members.
participation. We need not show that the first type of belief has greater effect than the others in reducing self-doubt in order to infer that it would be internalized, only that it is consistent with them.

Any set of beliefs about the public benefits arising from state control will involve at some level causal beliefs regarding the effects of state policies on utility-relevant goods. Clearly, in order to reduce self-doubt such beliefs must link positive marginal public benefits to group control of the state, and in turn to some kind of change in state policy rather than retention of the status quo. Given the multiple sets of policy changes that can be envisioned as having public benefits, the attention of members will focus on the most prominent available ideology that contains such beliefs and does not contradict other doubt-reducing belief changes.

Even without considering concrete cases, there are few conceivable reasons why any ideology regarding the public benefits of changes in state policy would logically constrain a member from adopting doubt-reducing beliefs about the importance of participation to group success or private benefits from participation. There are two ways in which participation can increase the probability of success: through the member’s own contribution to group success or through his or her role in promoting the participation of other members. Both are largely linked to the member’s beliefs about his or her own personal qualities, not to beliefs about the effects of group success per se. Beliefs about the effect of participation on private benefits would be linked to the extent to which participants would receive positive discrimination from other group members and perhaps to some intrinsic pleasure gained from being a participant.

The theory of ideology derived from the identity formation model implies that members of an opposition group seeking state power will share a common identity-related need to internalize beliefs in the benefits of state policy change and that the specific set of beliefs adopted by members will tend to converge around the most prominent available ideology that contains such beliefs. Such an implication in turn reduces the levels-of-analysis problem of analyzing the policies followed by such groups after they obtain power. Although members of such groups may retain disagreements about those elements of policy that relate to their own personal benefit, it will tend to promote unity with regard to those elements of policy that relate to collectively desired outcomes such as economic growth. Hence, at least for those elements of policy, they will increase the extent to which the state can be analyzed as a unitary entity without empirically problematic reductionism.

Anticolonial Action, Identity Formation and Economic Ideology
The theory as applied here focuses on former Western colonies, specifically those that gained independence in the post-World War II
Former colonies are perhaps the best large set of cases to which to apply this theory because they involve conflict between two clearly defined groups (the colonial rulers and the indigenous pro-independence elite), in which the group with status quo control was clearly identified with a particular economic ideology. Furthermore, these conflicts have always resulted in post-independence governments that were dominated by those indigenous elites who had participated in pro-independence activities, which allows us to compare resulting economic policies in a systematic manner. These factors free the analyst to a large extent from making fine-grained determinations about the boundaries of the groups involved in a political conflict, the ideologies associated with each, or how the conflict was resolved with regard to each group. At the same time, it provides a large set of cases for analysis.

With regard to these anticolonial struggles, the theory predicts that there will be a pronounced tendency for pro-independence elites to internalize oppositional ideologies. The interpretation of oppositional ideology in this application of the theory is made in an extremely straightforward and rather simplistic manner. It is posited that Western colonial powers both promoted a capitalist ideology and suppressed socialism and that this caused pro-independence elites to internalize socialism to bolster their decisions to engage in anti-colonial activity.

The reasons why socialist ideology would be internalized follow rather straightforwardly from the explication of the identity formation model in the previous section. It is clear that participation in the pro-independence struggle involved considerable risks for those involved. The indigenous elites who initially engaged in such struggles were necessarily those that were educated and skilled enough to carry out the necessary organizational activities. As such, they were usually already on the highest rung available to indigenous peoples within the colonial hierarchy.

Even when pro-independence activity held out the possibility of even higher positions and compensation than they currently held, there

---

89 Countries that gained independence considerably earlier, such as those in Latin America, will not be studied as ex-colonies because, from the viewpoint of the theory, the effects of their colonial experience on the ideology of state elites will have been weakened over time to an extent it would not have any direct significant effect.

90 This was so even in the case of most pro-Western post-independence governments. Houphouet-Boigny of the Ivory Coast, though often considered the paradigmatic pro-Western leader, just basically escaped arrest for his nationalist activities against French rule. Many of his colleagues in the African Democratic Rally, who later came to dominate post-independence government did spend considerable time in French prisons. The same was for Yew and the People’s Action Party, despite the fact that he was at least during the middle part of his career considered among the most pro-Western of East Asian leaders. Likewise, the travails of Kenyatta with the British colonial administration are well-known.

91 Wallerstein for instance characterizes African pro-independence parties as “nationalist movements with populist undertones, led by emerging middle-class, largely urban, elements with strong grievances against the colonial administration.” See Wallerstein 1967, 500.
Choosing an Identity

was clearly much more to be lost in the short term than could be gained. Indeed, even a small level of participation in such activity usually meant the loss of their official positions within the colonial hierarchy, not to mention possible imprisonment or physical harm. Hence participation was an all or nothing affair, with certain current loss balanced against uncertain future benefits. Although the decades after World War II brought about circumstances increasingly favorable to the end of colonization, there was no guarantee of timing nor the positions that pro-independence elites would hold in the postindependence order. Given the assumption that these elites were rational, we must infer that participation in such activity held higher levels of expected utility than collaboration. However, it is reasonable to infer that for most participants, given their existing beliefs, there were some possible circumstances under which the overall benefits would not exceed the costs.

Application of the identity formation model implies that the elites would search for a suitable ideology for reducing the resulting doubts. Socialist ideology was suitable because it met all the criteria for internalization. First, it provided a clear alternative to the capitalist ideology that was perceived to be underlying the status quo colonial economic policies. It promised increased economic productivity and reduced fears of the harm that separation from the colonial power would cause. Therefore, internalizing the ideology increased the perceived overall benefits that could accrue from indigenous control over the state, as long as socialist economic policies were included in postindependence policy plans.

Second, the ideology implied that the benefits of state control would be distributed to all members of society, not simply to those who happened to attain the top positions of power in any postindependence regime. Therefore it tended to eliminate the sources of self-doubt, which were attached to the least favorable set of possible circumstances for each member of the group. It is important to note that the suitability of socialism in this regard was not due to any promised egalitarianism but rather to the promise that sufficient benefits would accrue to all members of the postindependence order. The identity formation model implies that self-doubt is not generated by disproportionate benefits per se, but rather the perception of some individuals that they might, given an unfavorable state of the environment, receive insufficient benefits to justify their actions.

Finally, the socialist ideology was known to be available to all members of the group, and was clearly the most prominent set of beliefs on the publicly available ideational menu that satisfied the previous two criteria. Although there were undoubtedly other beliefs that elites could adopt that would emphasize the benefits of independence, the suppression of socialism and the resulting polarization of economic ideologies into two opposing alternatives provided pro-independence elites with a ready-made alternative which fulfilled their identity needs. Adopting different ideologies, even if they served the same needs for each individual, would have led to the expectation of future conflicts of interest, which in turn would have reduced the perceived benefits.
of collective action. Hence, as noted in the section on ideology in chapter 3, the model implies that beliefs will converge around the most prominent available ideology that met all members' identity needs.

The decision by a member of the indigenous elite to move from collaboration to conflict can lead to intellectual torment over all aspects of colonial ideology. Ideological change on a variety of fronts, via bolstering, can in turn lead to an “avalanche” effect which can completely alter the outlook of the former collaborator. The collaborationist indigenous elite generally exhibits an almost slavish desire to assimilate, one that includes an internalization of most aspects of colonial economic, political, and social ideology. This colonial syndrome of psychological dependence can itself be explained as a bolstering effect, because such ideologies will increase the perceived benefits that result from engaging in collaborationist activity. Once the (usually irrevocable) decision is made to engage in conflictual activity, however, the bolstering process begins to work in the opposite direction and the extent to which internalized ideologies change can be like a kind of religious conversion. As Fanon writes, the educated, middle-class colonial subject “initially flings himself upon the imposed culture” but later rejects it with equal passion.

Discussion of perceived benefits, of course, requires some assumptions about what exactly it is these elites are maximizing. Possible assumptions from the rational choice theories listed in the earlier review include personal wealth, national economic development, the welfare of one’s ethnic or national group (predatory, developmentalist, and nationalist preferences, respectively) and the ability to remain in power if one achieves it (conventional “Downsian” preferences). Without attempting to select a single maximand for preeminence, this indicates that the ideology that is most doubt-reducing is that which allows nationalist elites to believe that a single set of policies is optimal for maximizing all of them, i.e., that “all good things go together.” One characteristic of the major economic ideologies, such as capitalism or Marxism, is that each claims to provide the best strategy for achieving all the major plausible national economical goals such as growth, autonomy, and stability (as well as major political goals, such as democracy). It would also be reasonable and bolstering for such elites to believe that, ceteris paribus, policies that maximize national economic goals would generate a greater surplus that could be appropriated by state elites after independence, and would increase both their popularity and their security in power.

Because they affect the perceived costs and benefits of different economic policies, these ideologies will in turn create an impetus for

---

92 See Mannoni [1950] 1964; part I. See also critique in Fanon 1967, chap. 4.
93 Fanon 1976, 51.
95 Neither socialism nor capitalist policy can be clearly shown empirically to advantage or disadvantage accumulation of personal fortunes by political elites, as attested to the great wealth amassed while in office by Sukarno and Nkrumah on the one hand and Mobutu and Marcos on the other.
the elite groups to implement particular types of economic policies in the postindependence era, in a way that goes beyond that which can be explained purely by variations in structural economic and political factors. However, policy tendencies in either type of country will tend to decline over time, as first-generation nationalist elites are replaced with second-generation elites that have no experience in the anticolonial struggle and hence do not have the same ideological predispositions.

Assumptional Issues
Concretely, colony is defined as a country where ultimate state authority (i.e., the ability to make final decisions on policy matters) has been predominantly controlled by actors outside the national territory for a period of at least five years. The major Western colonial powers (Britain, France, Holland, Portugal and Belgium) are assumed to all have been associated with capitalist ideologies, although later attempts will be made to investigate possible differences in the effects of British and French colonialism. Japan, however, is not be seen as associated with a clear economic ideology, and hence no predictions are made about the effects of colonial experiences on the ideologies of nationalist elites in its ex-colonies.

These definitions and auxiliary assumptions immediately bring up obvious questions of reductionism and oversimplification. First of all, it can be questioned whether all the colonies, as defined here, underwent the same types of experiences under foreign domination, given the differences between British indirect rule and French centralization.96 Second, there were undoubtably differences in the types of economic policies followed by each of the Western colonial powers. Moreover, none of them followed anything that approached “pure” capitalist nonintervention, and all placed numerous restrictions and regulations on agricultural production within their colonies.97 Likewise, some might argue that Japan ought to be associated with a clear economic ideology, that based on a dominant alliance between a strong bureaucracy and private conglomerates.

With regard to whether ideologies and economic policies found in Western colonies should be called capitalist, the important thing is not whether they met some free-market ideal but whether which side they were on within the bipolar economic ideology spectrum. And as far as nationalist elites were concerned, they were clearly on the capitalist side. The link between capitalism and (undifferentiated) Western colonialism has been the central analytical focus of both the Leninist theory of imperialism and prominent theories of underdevelopment,98 and a cursory glance at the writings of prominent nationalist elites

---

96 For contrasts between different systems of colonial rule, see Smith 1981c and Kahler 1984.
97 See for instance Young 1982, 189.
98 Most notably Baran 1957 and Frank 1966.
under various Western colonial powers shows a strong tendency for them to agree with this assessment.\textsuperscript{99}

For instance, Nkrumah accepts a fairly straightforward interpretation of the Leninist theory, defining the enemy as “international finance capital under various external and internal forms of exploitation, imperialism, and capitalism.”\textsuperscript{100} Amilcar Cabral similarly discusses “imperialism, which everything goes to show is really the last stage in the evolution of capitalism.”\textsuperscript{101} Nehru talks about “the capitalist order which had inevitably developed into vast imperialisms, which . . . swallowed the colonial world.”\textsuperscript{102} Senghor writes, “Negro-African society was intellectually, technically, politically and therefore economically dominated by European capitalism—French capitalism, in our case.”\textsuperscript{103} The political manifesto of the Sarekat Islam movement, the dominant Indonesian nationalist group, decries the evils of “sinful capitalism.”\textsuperscript{104}

The related question of whether Japanese colonialism can clearly be associated with an economic ideology, i.e., the model of bureaucracy and private conglomerate-led development that Johnson has dubbed the “capitalist development state,”\textsuperscript{105} is pertinent because the perceived tendency of South Korea and Taiwan (though obviously not North Korea) to follow this development path. This in turn would seem to point against the prediction of oppositional ideology formation. However, the incorporation of such a mode of development into a clear development ideology has a very recent vintage, and is the product more of Western scholars\textsuperscript{106} than of the Japanese themselves. Hence, the Korean and Chinese nationalist elites fighting against the Japanese in the 1930s and 1940s could hardly have been aware of such an ideology. More importantly, the Japanese state itself during that period was dominated by an agrarian-based military clique that may have tolerated the zaibatsu conglomerates but viewed them with great suspicion and attempted to stem their power in favor of a more centrally-controlled economy. Given this historical context, it would be implausible to suggest that the adoption of an anti-zaibatsu ideology would have had a significant bolstering effect for nationalist elites, and it is not that surprising from the standpoint of the theory that this particular feature of the colonial economy might survive independence.\textsuperscript{107}

\textsuperscript{99} For further discussion on this general point, see Sigmund 1972, introduction to the first edition, 13-15 and Anderson, Van der Meiden, and Young 1967, 193-94.
\textsuperscript{100} Nkrumah 1968, 1-2.
\textsuperscript{101} Cabral 1979, 127.
\textsuperscript{102} Nehru [1941] 1958, 392.
\textsuperscript{103} Senghor [1961] 1964, 96.
\textsuperscript{104} Cited in Dahm 1969, 35.
\textsuperscript{105} Johnson 1982
\textsuperscript{106} In addition to Johnson, see the other “revisionists” cited above.
\textsuperscript{107} In fact, the only clear ideology that was associated with Japanese colonialism was that of pan-Asianism under imperial rule, which was in turn manifested by the Greater East Asian Co-Prosperity Sphere and the desire to unify the “eight corners” of the earth. The rejection of pan-Asianism has been notably complete
Without a doubt, much finer distinctions can be made between countries and colonial experiences than is being made here, and some of these distinctions will have had significant consequences for state ideology and policy. However, incorporating such distinctions would make the theory considerably more complex and would divert attention from the central purpose, which is to show how the factors discussed here had systematic effects on internalized ideologies and in turn on economic policies. As noted earlier, the theory is not meant to be an all-encompassing explanation of ideology and policy but rather a systematic and theoretically informed exploration of one vivid pattern of linkages.

A Preliminary Assessment of Economic Ideology
Because it is impossible to directly measure internalized ideology, any preliminary attempt to assess the predictions of the theory with regard to ideology must be based upon examination of the ideologies that state elites espouse. An examination of the historical record does provide substantial evidence that the espoused ideologies of state elites in ex-colonies conform with the predictions made by the theory. The tendency of Third World countries, which are nearly all ex-colonies, to espouse socialist ideologies has long been noteworthy. As the novelist Soyinka has put it, there has been a well established tendency within the Third World of “excusing . . . leaders simply because they are on the ‘left’.” Furthermore, as the theory predicts, espousal of socialism in ex-Western colonies is closely associated with nationalism and the pursuit of independence. Young sees “the roots of African socialism in anticolonial nationalism,” and shows how the two were often closely linked in postcolonial ideology. In the newly-independent countries of the Third World, socialism has often been considered an integral part of the nationalist “political religion.” Bell notes, “for some individuals, anti-imperialism is equated with being socialist.” As Plamenatz puts it, “Socialism and hostility to empire often go together, especially in the early stages [of independence].” Or, as Josey writes, “Socialists within Japanese ex-colonies, as has the rejection of any major international political role for Japan.

108 See for instance the chapters containing excerpted writings by Third World leaders on “Indian Socialism,” “Arab Socialism,” and “African Socialism” in Sigmund 1972. There are no corresponding chapters on Indian, Arab or African capitalism. See also Andrain 1964; Anderson, Van der Mehden, and Young 1967, chap. 10. In Africa, Young has done a survey of the basic development strategies of 17 African countries, all of which are ex-Western colonies. Eight are classified as “Afro-Marxist,” six as “populist socialist,” and only three as “African capitalist.” See Young 1982, chap. 1, also table 6.1 on p. 299.


110 Apter 1963, 99. See also the introduction to Apter 1964.

111 Bell 1975, 150-51.

112 Plamenatz 1960, 130.
and nationalists generally—and in Asia a Socialist is first a nationalist—automatically opposed capitalism because they opposed colonialism with which it was associated."

Despite the frequent observation that anticolonial nationalism and socialism tend to go together in the Third World, there is a notable absence of theoretical explanations in the political development literature of why this is so. The theory outlined above provides such an explanation, showing how this linkage is rooted systematically in the psychological processes of bolstering, which is part of the more general doubt-reduction process, and in the nature of Western colonialism and the anticolonial struggle.

While the theory predicts that nationalist elites in ex-Western countries will reject capitalism and embrace a socialist ideology, this ideology need not conform with orthodox Marxism. In fact, many Third World leaders have taken pains to emphasize the indigenous nature of their economic ideologies. As Bell goes on to note, there is a great deal of diversity among countries that label themselves socialist. He states that “what ‘socialism’ means as a positive socioeconomic program, or as an ideology or doctrine has become hard to define.” Rather, it is “an important symbol, even in this negative sense, against capitalism.” Or as Anderson, Van der Mehden and Young put it, “developmental socialism is anything but a well-delineated doctrine. Rather, the socialists of the developing nations endorse a bewildering variety of beliefs, theories and action programs.”

However, while it is true that the socialism is a very broad label, it is also true that all Third World socialisms tend to emphasize the benefits of state economic intervention. As Sigmund notes,

> there are some common themes in the discussion of socialism in developing countries. There is a belief in the necessity of government action to stimulate economic development and a corresponding rejection of laissez-faire liberalism. There is an increasing tendency to nationalization or divestment on a fixed schedule. . . . There is, most important, a belief that rational planning can result in the elimination of waste, inefficiency, and selfishness, both domestically and in international economic relations.

Despite the differences between Third World socialisms and orthodox Marxism, he notes that “because of the involvement of capitalism in the colonial enterprise, the criticisms of capitalism often

---

114 Josey 1957, 2.
115 See also chapters in Eisenstadt and Azmon 1975.
116 Most notably, the development of “African Socialism.” For an overview, see Friedland and Rosberg 1964. For writings by perhaps the most prominent exponent, see Nyerere 1968.
117 Bell 1975, 151
118 Anderson, Van der Mehden, and Young 1967, 176.
119 Sigmund 1972, introduction to the second edition, 49.
seem considerably harsher than those of Communism. Even native capitalists are linked to the colonial rulers in nationalist theory.”

The theory suggests that the decision to adopt socialist ideologies was not a result of cold objective evaluation of economic and political conditions but rather a motivated psychological reaction to the policies of the colonial powers. As Ugandan president Museveni has said about African ideologies: “Our heads were in Europe. We saw Africa in terms of European conflicts.”

This psychologically driven process of opposition ideology formation is not restricted to the economic sphere. An analogous phenomenon can be found in the cultural sphere in the formation of the ideology of “negritude,” which defines itself not on the basis of traditional African distinctions but in opposition to the colonially-imposed culture. As Senghor writes, “Paradoxically, it was the French who first forced us to seek its essence, and who then showed us where it lay... when they forced their policy of assimilation and thus deepened our despair.” While the Western colonial powers were unable to implant their ideologies in the minds of the indigenous elite, it is clear that they were able to establish the dimensions along which ideological conflict would take place. As Plamenatz puts it, “How tempting it is to say to alien, and often arrogant, intruders, ‘Here is something as good as anything you have to offer, and better; if you try to prevent our having it, we must push you out of the way’.”

The theory would predict, however, that not all aspects of colonial beliefs and values will eventually be rejected by those engaging in nationalist activity. While it is psychologically bolstering to internalize ideologies that emphasize the shortcomings of colonial policies, it will also be bolstering to emphasize the advantages of those products of colonial rule that have been denied to indigenous peoples and can be seized along with state power. This can explain why states in Third World ex-colonies have placed special value on building Western-style institutions of higher education for indigenous peoples and on copying the titles and lifestyles of Western government officials. It also explains why Third World state elites, even when they profess strong anti-Western sentiments, vigorously seek to be included in Western-originated international institutions, as long as interactions there are on a level of equality and not the hierarchical transactions that were experienced under colonialism.

Finally, although they were not colonies in the conventional sense, it is interesting to apply the theory to the former satellites of the Soviet Union, which after all faced much the same type of foreign domination.

---

121 Museveni 1991.
123 Plamenatz 1960, 29.
124 For a discussion of this tendency, see Tan and Mingat 1992.
125 Abernethy 1988.
as many other types of colonies. Because of the relative recentness of their “independence,” data on policy are more difficult to come by, particularly with regard to change over time after independence. Moreover, there is a much greater variation than with Western colonies with regard to whether initial control of the post-independence state was by nationalists or former collaborators, something that can be attributed to the fact that the higher levels of repression in the ex-Soviet satellites inhibited the emergence of a large group of viable nationalist leaders. However, rejection of socialism in virtually every country and the almost mystic reverence initially paid to capitalism in countries where nationalists came to power seem to concur very strongly with the predictions of the model. These views seem to continue to be held by nationalist elites, even in light of widespread public reaction against the hardships of the radical economic transition.

**Alternative Explanations for Economic Ideology**

At this point, it would be useful to discuss some alternative explanations that might be made for the general phenomenon in which nationalist elites espouse economic ideologies that oppose those of their colonial power.

First of all, it might be argued that such an oppositional ideology was held all along by nationalist elites and that in fact it may have been the *cause* rather than the effect of political conflict. It is difficult, of course, to trace the path of intellectual causation in each set of state elites, not the least because members of such elites, being politicians, are hesitant to admit that they have ever failed to hold the ideas that they hold currently.

However, it can be noted that, within pro-independence movements in Western colonies, socialism came to be a dominant ideological force rather late in the game. As mentioned earlier, all prominent anticolonial parties had their beginnings in bourgeois nationalism, i.e., in the grievances of highly-educated and relatively assimilated members of the indigenous middle class. Hence, anticapitalism, even if it was mentioned, was certainly not stated as the major justification for opposing Western colonialism. In some cases, prosocialist factions were actually relatively weak to begin with and only gradually built up their power and prominence within the nationalist movement, eventually becoming dominant.

Another alternative explanation would accept the endogeneity of ideology formation but ascribe it to different causes, doing so by separating espoused ideology from internalized beliefs. It would argue that nationalist elite groups had an incentive to espouse oppositional economic ideologies because this was a useful device for attracting mass support for political independence. This view of ideology,

---

126 Anderson, Van der Mehden, and Young 1967, 199.

Choosing an Identity

however requires that the “masses” be viewed as passive vessels who are uncritically accepting of arguments put forward by elites. Furthermore, as noted earlier, the problem with purely instrumental explanations of ideology is that political actors in a world where ideology is purely instrumental will not take seriously any of the ideological statements of other actors, hence making it unclear why ideology would ever be espoused in the first place. Finally, a key argument of the theory is that oppositional economic ideologies, because they were internalized, significantly affected policies that were followed in the postcolonial era, a proposition that will be tested in the following sections.

A somewhat different explanation based on the instrumental use of ideology would argue that elite groups had an incentive to espouse socialism in order to gain financial support from international sponsors such as the Soviet Union and China. There are a number of problems with this explanation, not the least of which is the fact that little sustained support was actually forthcoming, even for the most dedicated of professed socialist ex-colonies. In fact, the overall amount of resources flowing from the socialist superpowers to ex-colonies has been much lower than the resources flowing from Western countries. Furthermore, a purely instrumental explanation would not explain why ex-colonies would not seek aid from their even wealthier ex-colonial powers, nor why the socialist powers would care about the domestic political arrangements of their potential allies. Finally, this explanation does not fit easily with the fact that the leaders of ex-colonies generally took great pains to distinguish their own brand of socialism from the Soviet or Chinese kind and often persecuted their own domestic communist parties.

A final explanation would accept the internalization of ideology but would ascribe it to different causes. It would focus in particular on intellectual fashions and the supposed wide acceptance of the superiority of interventionist economic ideologies among both Western and Third World elites during the 1950s and 1960s. However, if the worldwide intellectual zeitgeist determined ideological choice, then it is difficult to explain why states in Western industrialized countries were not as quick as those in their ex-colonies to adopt socialist ideologies. Furthermore, it does not explain why the surviving first-generation leaders of ex-colonies (e.g., Kaunda and Nyerere) have largely refused to abandon their adherence to socialism, even as it has been largely rejected in most of the West.

Finally, it is difficult to use an intellectual fashion argument as a basis for predictive theory, since it provides no way determine to the direction in which fashion will move in the future, regardless of whether this change is due to learning or other factors. However, it is useful to take into account global intellectual changes, even after the fact, and hence an effort is made to do so in the empirical test that follows.

129 Emerson 1962, 376-77.
Economic Policy in the Post-Independence Era
Because the theory presented here combines a coherence model of ideological preference and belief formation with a rational actor model of action, ideology is an intermediate dependent variable that is influenced by independent variables (the political experiences of elite groups) and in turn influences a further dependent variable (the policy choices of these groups when they gain control of the state). Inferring the predicted effects of internalized ideologies on the policy choices of state elites is straightforward, though it requires some analysis of how the internalized ideologies of individual nationalist elites will influence the collective policy decisions of these elites when it comes into control of a state.

One clear implication of the argument given here is that joint participation in proindependence activities will promote homogeneity among nationalist elites with regard to economic ideology. Because they are assumed to be rational, shared ideological beliefs that emphasize the collective and personal benefits of a certain set of policies will make a group of elites more likely to favor such policies than a group of elites that have not internalized such beliefs. Of course, there may be significant collective action problems that stand between shared desires for certain policies and the actual implementation of such policies. However, it is reasonable to assume that they are not so strong as to eliminate some sort of positive connection between policy preferences and implementation.

Hence, it can be inferred that if a state is dominated by elites that have internalized an ideology emphasizing the economic benefits of state intervention, it will, ceteris paribus implement more interventionist policies than a state that has not internalized such an ideology. Hence, because states in newly independent ex-Western colonies were dominated by elites that had engaged in proindependence activities, the theory will generate the hypothesis that such states will be more likely to implement interventionist policies than other states, even after taking into account differences in economic structural factors.

Another hypothesis generated by this model is that the effects of the oppositional ideology formation on economic policy in ex-colonies should decline over time after independence, as those of the initial generation of state elites retire or die and new generations of elites whose members have no direct experience in anticolonial political activity move into power.130 Having not participated in the anticolonial struggle, there will be no psychological reason for them to have internalized the oppositional ideology, and hence they will have a different view of the benefits and costs of interventionist policies.

---

130 This analysis, of course requires some fudging of the unitary model of the state. However, it can be retain formally by viewing this generational turnover as a structural attribute of the state rather than a disaggregation of it.
This in turn will lead to decline in the distinctiveness of the policies of ex-colonies over time.\textsuperscript{131}

\textbf{A Statistical Analysis of Economic Intervention}

An attempt was made to test these implications empirically, using pooled cross-sectional and time-series country data from taken from three different datasets, the Penn World Tables (PWT)\textsuperscript{132} as well as the International Monetary Fund’s International Financial Statistics (IFS)\textsuperscript{133} and Government Finance Statistics (GFS).\textsuperscript{134}

The dependent variables, which measure levels of government intervention, were of two types, corresponding to the two most common ways in which intervention is conceptualized. The first is the size of government with respect to the total size of the economy, or government spending versus total private spending in areas that would be outside its purview in a pure market economy. The second is the extent to which intervention (e.g., tariffs, quotas, licenses, subsidies and currency controls) distorts the overall prices of goods from their free-market value.\textsuperscript{135}

The first type relates to the level of government size relative to the total size of an economy. This is a quite intuitive measure of levels of state intervention, and coincides with the notion of “big government.” However, there are a number of ways in which such spending can be measured, and three different measures will be tested. The first is simply government expenditure as a percentage of gross domestic product. One possible problem with this measure is that it does not adequately make a distinction between types of spending that might be consistent with a capitalist ideology (such as defense, education and certain types of social welfare) and those types that clearly are not. Direct government support or ownership of economic enterprises, however, is the one area where socialists and capitalists clearly part company. Hence, the second measure of government size is spending purely on “economic services,” which:

comprises expenditure associated with the regulation, support, and more efficient operation of business; economic development; redress of regional imbalances; and creation of employment opportunities. Research, trade

\textsuperscript{131} In fact, if there is considerable political conflict between the first generation of anti colonial elites and a group of newer challengers to power, the theory would predict that a “counter-oppositional” ideology to form around the challengers, and a sort of “cycling” may occur. For a discussion of this sort of phenomenon, See Coleman 1965, 27.

\textsuperscript{132} For a detailed description of the original version of the dataset, see Summers, Kravis, and Heston 1980. For a description of a more recent version, see Summers and Heston 1991. The version used here is Mark 5.6, which has data up to 1992. This is available via anonymous ftp from nber.harvard.edu.

\textsuperscript{133} The 1992 version, with data from 1948-91, which is available from ICPSR.

\textsuperscript{134} The 1995 version, with data from 1971-94, which is available from ICPSR.

\textsuperscript{135} For broad discussions of these measures, see Woo 1990 and Perkins 1991.
Ideology Formation and Policy Choice in Ex-Colonies

promotion, geological surveys, and inspection and regulation of particular industry groups are among the activities included.\textsuperscript{136}

Finally, either measure of government size may distort the government’s role in producing goods and services because it reflects the amount spent rather than the actual value of the goods and services at international prices. These may be different because of additional policies that intervene in the price mechanism. Hence, the third measure of government size will be the government goods and services share of total production, measured in terms of constant international prices.\textsuperscript{137}

The other type of dependent variable is a measure of aggregate price distortion in each country’s economy. This variable represents the extent to which state intervention causes the domestic price of goods to diverge from the international price, and it is influenced by a variety of government economic policies, including restrictions on trade in goods and services, on currency exchange, and on domestic production. The indirect measure of distortion used here is the gap between the domestic prices for a representative basket of goods and international prices at official exchange rates measured in percentage terms.\textsuperscript{138} It is generally assumed that higher levels of intervention will lead to higher aggregate prices rather than lower ones, because intervention will generate inefficiencies.\textsuperscript{139}

The first set of independent variables to be regressed against the above dependent variables are dichotomous dummy variables that represent the colonial histories of various countries. One variable represents whether or not a country is a former Western colony that gained independence in the twentieth century, taking the values zero if false, one if true. While the theory does not make any clear predictions about the effects of specific subtypes of Western colonialism, two other analogous variables are included to test whether a history of British

\textsuperscript{136} World Bank 1991, 279.

\textsuperscript{137} Overall spending levels were obtained from the International Financial Statistics dataset, while economic services spending levels were obtained from the Government Finance Statistics. They were divided by measures of GDP, obtained from the same datasets. Price-corrected goods and services as a share of GDP were obtained from the Penn World Tables.

\textsuperscript{138} This was obtained from the Penn World Tables. It has been used as a measure of government intervention in Dollar 1992, and also adopted in the 1991 issue of the World Development Report. See World Bank 1991, chapters 4 and 5. A discussion of the justification for using aggregate relative prices as a measure of government-induced distortion can be found in Dollar 1992, 525-27.

\textsuperscript{139} One way of testing this is to check exchange rate premiums (the gap between official and black market exchange rates. An official exchange rate which assigns a positive premium to local currency indicates distortions which inflate aggregate local prices relative to international prices, while a negative premium indicates distortions which deflate local prices. A premium of zero indicates nondistortive policies. See World Bank 1991, 82 for a discussion. Examination of 1990 estimates in the International Currency Yearbook shows no cases where a negative premium occurs. See International Currency Analysis 1993
Choosing an Identity

or French colonialism had any significant impact separate from that of undifferentiated Western colonialism.\footnote{A total of 74 countries in the three datasets were classified as ex-Western colonies. They were (with year of independence and colonial power in parentheses, B=British, F=French, O=Other): Algeria (62 F), Angola (75 O), Bahamas (72 B), Bangladesh (47 B), Barbados (66 B), Belize (81 B), Benin (60 F), Bhutan (49 B), Botswana (66 B), Burkina Faso (60 F), Burundi (62 B), Cameroon (60 O), Cape Verde (75 O), Central African Republic (60 F), Chad (60 F), Comoros (75 F), Congo-Brazzaville (60 F), Djibouti (78 F), Fiji (70 B), Gabon (60 F), Gambia (65 B), Ghana (57 B), Grenada (74 B), Guinea (58 F), Guinea-Bissau (74 O), Guyana (66 B), India (47 B), Indonesia (49 O), Ivory Coast (60 F), Jamaica (62 B), Kenya (63 B), Laos (53 F), Lesotho (66 B), Madagascar (60 F), Malawi (64 B), Malaysia (57 B), Mali (60 F), Malta (64 B), Mauritania (60 F), Mauritius (68 B), Morocco (56 F), Mozambique (75 O), Myanmar (48 B), Niger (60 F), Nigeria (60 B), Pakistan (47 B), Papua New Guinea (73 B), Philippines (46 O), Reunion (90 F), Rwanda (61 O), Saint Lucia (79 B), Saint Vincent and Grenadines (79 B), Senegal (60 F), Seychelles (76 B), Sierra Leone (61 B), Singapore (59 B), Solomon Islands (78 B), Somalia (60 O), Sri Lanka (48 B), Sudan (56 B), Surinam (75 O), Swaziland (68 B), Syria (46 F), Tanzania (61 B), Togo (60 F), Tonga (70 B), Trinidad and Tobago (62 B), Tunisia (56 F), Uganda (62 B), Vanuatu (80 O), Western Samoa (62 O), Zaïre (60 O), Zambia (64 B), Zimbabwe (60 O). Former British protectorates where indigenous rulers retained a measure of formal political control for long periods of time (e.g., Egypt, Iraq, Jordan, Nepal, and the gulf states), were not considered as ex-colonies, although their inclusion did not significantly change the results of regressions. Among the former colonies not covered in any of the datasets are Cambodia, Equatorial Guinea, Lebanon, Libya, and Vietnam. The Latin America category included Antigua Barbuda, Argentina, the Bahamas, Barbados, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, the Dominican Republic, Ecuador, El Salvador, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Surinam, Trinidad and Tobago, Uruguay, and Venezuela. The African category included Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, the Central African Republic, Chad, Comoros, Congo-Brazzaville, Equatorial Guinea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, the Ivory Coast, Kenya, Lesotho, Liberia, Madagascar, Malawi, the Maldives, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Reunion, Rwanda, Sao Tome Principe, Senegal, the Seychelles, Sierra Leone, Somalia, South Africa, Sudan, Swaziland, Tanzania, Togo, Uganda, Zaïre, Zambia, and Zimbabwe. The East Asian category included Cambodia, China, Hong Kong, Indonesia, Japan, North Korea, South Korea, South Korea, Laos, Malaysia, Mongolia, Myanmar, the Philippines, Singapore,}

Included also are variables that detect whether the presence of certain groups of countries within the ex-Western colony category or outside it may be spuriously causing a connection between a history of Western colonialism and state intervention. As was discussed earlier, elites in Latin American countries that gained independence in the 19th century are not expected to have any systematic caused bias toward interventionist ideology. Hence, other than a few former British Caribbean possessions, Latin American countries are not placed in the ex-Western colony category. However, a dummy variable for Latin American countries is used to control for them as a group and measure their independent impact on the results. Likewise, dummy variables are included for Sub-Saharan African countries (all of which were ex-Western colonies) and East Asian countries.\footnote{The Latin America category included Antigua Barbuda, Argentina, the Bahamas, Barbados, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, the Dominican Republic, Ecuador, El Salvador, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Surinam, Trinidad and Tobago, Uruguay, and Venezuela. The African category included Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, the Central African Republic, Chad, Comoros, Congo-Brazzaville, Equatorial Guinea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, the Ivory Coast, Kenya, Lesotho, Liberia, Madagascar, Malawi, the Maldives, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Reunion, Rwanda, Sao Tome Principe, Senegal, the Seychelles, Sierra Leone, Somalia, South Africa, Sudan, Swaziland, Tanzania, Togo, Uganda, Zaïre, Zambia, and Zimbabwe. The East Asian category included Cambodia, China, Hong Kong, Indonesia, Japan, North Korea, South Korea, South Korea, Laos, Malaysia, Mongolia, Myanmar, the Philippines, Singapore,}
dummy variables can be used to control for the systematic independent effect of geographical and cultural factors on policy.

A final variable in the set measures the number of years that have passed since independence for former colonies and takes a value of zero for countries that were not ever Western colonies. This variable, in addition to testing the hypothesis that levels of intervention will revert to the mean with respect to time after independence, is also relevant to testing the alternative (and opposing) theory that country age will have a positive effect on levels of economic intervention.\footnote{The achievement of independence can be likened to the boundary change or other disruptive events which Olson uses to mark the beginning of the formation of distributive coalitions. See Olson 1982}

The second set of independent variables to be regressed comprises each country’s economic structural characteristics. These include total and per-capita measures of gross domestic product (in constant dollars) and of land area as well as total population. One major effect of these variables is to purge the aggregate price variables of influences other than state-induced distortion, such as the effect of relative capital, land and labor endowments on the prices of nontradeable goods.\footnote{Among other authors who have used aggregate prices as a measure of intervention, Dollar purges for labor and land costs by including independent variables for per capita GDP and population density. See Dollar 1992, 526-28. The World Bank uses a slightly different methodology, with variables for urbanization, land area and population. See World Bank 1991, 163.}

The variables also have the effect of accounting for some of the most prominent existing explanations for variations in government intervention. The percapita measures of GDP can act as a way of incorporating wealth-related influences on government size, à la Wagner’s Law as well as the effects of wealth on price-distorting protectionism. Furthermore, the measures of total GDP, land and population can be used as ways of incorporating the effects of country size on both government size and price distortion.\footnote{Data for GDP and populations variables were obtained from the same datasets as the dependent variables against which they were regressed. IMF GDP data was in local currency, so figures in constant dollars were obtained by dividing each by the current local currency/dollar exchange rate, then by the current GDP deflator for the U.S. (1985 dollars = 1 for International Financial Statistics and 1990 dollars = 1 for Government Finance Statistics). Data for land area were obtained primarily from the World Development Report.}

The third set of independent variables are dummy variables representing every year for which datapoints exist within each dataset, each taking the value one if the observation is for that year, zero if not. These variables will account for the effects of global historical changes, such as changes in intellectual fashion, economic cycles and shocks, and shifts in the superpower balance of power, on variations in state economic intervention over time.

\textit{Results of Regression Analysis}

Two models were tested for each of the four dependent variables: overall government spending/GDP, government spending on business

\textit{Taiwan, Thailand, and Vietnam.}
Choosing an Identity

production/GDP, price-corrected government spending/price-corrected GDP, and aggregate price levels. The first model included the five economic structural variables, the year by year dummy variables, and the three variables for colonial experiences as independent variables. The second model included all these variables as well as the variable for years since independence. When each dataset was reduced to those observations which included complete data for these models, they consisted of time series with varying starting points and lengths for different countries. In order to provide reasonably stable estimates of cross sectional heteroscedacity, observations for countries whose times series were shorter than 10 years were deleted. However, no attempt was made to convert the datasets into panels, because this would have resulted in deletion of most of the data for the IMF datasets, and an unrepresentative sample biased toward industrialized countries. Hence, the basic models were:

\[
\text{Dependent Variable} = \beta_1(\text{ex - Western Colony}) + \\
\beta_2(\text{ex - French Colony}) + \beta_3(\text{ex - British Colony}) + \\
\beta_4(\text{Africa}) + \beta_5(\text{Latin America}) + \\
\beta_6(\text{East Asia}) + [\beta_7(\text{Years Since Independence}) + ] \\
\gamma_1(\text{GDP}) + \gamma_2(\text{GDP/Capita}) + \gamma_3(\text{Area}) + \\
\gamma_4(\text{Area/Capita}) + \gamma_5(\text{Population}) + \sum_t \zeta_t(\text{Year} = t)
\]

No intercept term was included because it was subsumed within the year dummy variables.

Generalized least squares was undertaken, with detected AR(1) autocorrelation within time series being corrected for all regressions via a Prais-Winsten transformation, with a single autocorrelation coefficient assumed for all time-series.\(^{145}\) Heteroscedacity across cross-sections and time-series was subsequently corrected for via the method of weighted least squares, with the estimated weights being based upon the reciprocal of the sample standard deviation of the residuals for each cross section and time-series. Each vector of observations was transformed by multiplying it by the estimated weights for its corresponding cross section and for its time series. The relatively large number of countries covered relative to the length of time series made it difficult to compute any suitable estimate of contemporaneous cross-sectional correlation, but the year to year dummy variables pick up

\(^{145}\) See Prais and Winsten 1954. The coefficient was estimated as \( r = \frac{\sum_n \sum_{t \neq 1} e_{n,t} e_{n,t-1}}{\sum_n \sum_{t \neq 1} e_{n,t}^2}, \) where \( e \)'s are the residuals from an OLS regression over the dataset in question. Prais-Winsten estimates involved multiplying the vector of dependent and independent variables for first observation in each time series by \( \sqrt{\hat{r} - 1} \) and for subsequent observations subtracting from each variable in each vector \( r \) multiplied by the variable’s single-period lagged value.
global contemporaneous effects while the ex-colony dummy variables
pick up effects that relate their specific groups.

The results of these regressions are shown in Table 1. The major
result is that the estimated coefficients of the ex-Western colony dummy
variable have the positive (correct) sign and are highly significant
statistically across all seven of the eight regressions, including four out
of four when years from independence is included as an independent
variable. This supports the hypothesis that states in ex-Western
colonies will be more interventionist in their economies than those in
other countries even when structural, geographical and year to year
control variables are taken into account.

The ex-British colony variable coefficient is generally negative
across the regressions (except for the price-corrected government
spending regression), while the ex-French colony coefficient does not
show any clear pattern. However, only the ex-British colony coefficient
for uncorrected spending and spending for economic services are
statistically significant when the years since independence variable
is included. More importantly, the significance of the ex-Western
colony variable despite the presence of more specific ex-colony variables
implies that the effects of Western colonialism on levels of economic
intervention are general and not the result of the colonial policies
of a particular Western country, notwithstanding differences between
Britain, France, and other colonial powers.

None of the three geographical variables is significant across more
than half of the regressions. Furthermore, the Latin America and
East Asia dummy variables do not show any clear pattern in the
signs of their coefficients over the various regressions, indicating that
geography in these cases has no clear bearing on the tendency of states
to intervene. The Africa variable, on the other hand, does show a
pattern of positive coefficients, but they are only significant in two
of the eight regressions. Overall, no clear effect of geography can
be detected on levels of state economic intervention. The continuing
significance of the ex-Western colony variable with the dummy variable
included indicates that whatever the independent effect of geography,
colonial factors are important in shaping policy choices. Furthermore, it
strongly implies that, although being an ex-Western colony is strongly
correlated with geography, the former rather than the latter is a more
convincing explanation for differences in economic policy.

The coefficient for years since independence is negative (the
expected sign) and significant for all four regressions in which it is
included. This supports the hypothesis that the level of government
intervention in ex-Western colonies will decline over time after
independence. By implication, it tends to contradict the alternative
hypothesis that the growth of distributional coalitions will tend to lead
to growth in government size as a function of a country’s age. It should
be emphasized that, because of the presence of year to year dummy
variables, these results cannot be interpreted as an artifact of global
debates in average levels of government intervention over time. Nor
can they be attributed to exogenous economic shocks occurring on a
global scale.
### Table 1. Estimates for Levels of Government Economic Intervention

<table>
<thead>
<tr>
<th></th>
<th>Government Spending(^\dagger)</th>
<th>Government Spending on Economic Services(^\dagger)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex-Western colony(^a)</td>
<td>9.76*, 2.35</td>
<td>4.02*, 0.79</td>
</tr>
<tr>
<td></td>
<td>21.00*, 2.86</td>
<td>5.97*, 0.95</td>
</tr>
<tr>
<td>Ex-French colony(^a)</td>
<td>2.16, 3.23</td>
<td>-1.90, 1.09</td>
</tr>
<tr>
<td></td>
<td>1.90, 3.52</td>
<td>-1.94, 1.13</td>
</tr>
<tr>
<td>Ex-British colony(^a)</td>
<td>-1.22, 2.33</td>
<td>-2.37*, 0.76</td>
</tr>
<tr>
<td></td>
<td>-7.79*, 2.61</td>
<td>-2.93*, 0.79</td>
</tr>
<tr>
<td>Africa(^a)</td>
<td>2.66, 1.62</td>
<td>0.27, 0.47</td>
</tr>
<tr>
<td></td>
<td>1.52, 1.61</td>
<td>0.22, 0.47</td>
</tr>
<tr>
<td>Latin America(^a)</td>
<td>3.01*, 0.92</td>
<td>0.67, 0.31</td>
</tr>
<tr>
<td></td>
<td>0.80, 0.95</td>
<td>0.33, 0.32</td>
</tr>
<tr>
<td>East Asia(^a)</td>
<td>0.07, 1.56</td>
<td>-0.52, 0.47</td>
</tr>
<tr>
<td></td>
<td>-0.75, 1.58</td>
<td>-0.11, 0.48</td>
</tr>
<tr>
<td>Years since independence(^b)</td>
<td>—, 0.07</td>
<td>—, 0.09</td>
</tr>
<tr>
<td></td>
<td>-0.77*, 0.47</td>
<td>—, 0.11*</td>
</tr>
<tr>
<td></td>
<td>—, 0.09</td>
<td>—, 0.03</td>
</tr>
<tr>
<td>Total GDP(^c)</td>
<td>0.39*, 0.11</td>
<td>0.003, 0.02</td>
</tr>
<tr>
<td></td>
<td>0.28*, 0.10</td>
<td>-0.02, 0.02</td>
</tr>
<tr>
<td>Per capita GDP(^c)</td>
<td>0.02, 0.05</td>
<td>0.005, 0.01</td>
</tr>
<tr>
<td></td>
<td>-0.05, 0.05</td>
<td>0.006, 0.01</td>
</tr>
<tr>
<td>Land area(^c)</td>
<td>-0.03, 0.03</td>
<td>-0.01, 0.01</td>
</tr>
<tr>
<td></td>
<td>-0.03, 0.03</td>
<td>0.006, 0.006</td>
</tr>
<tr>
<td>Per capita land(^c)</td>
<td>1.86*, 0.37</td>
<td>0.48*, 0.10</td>
</tr>
<tr>
<td></td>
<td>1.57*, 0.36</td>
<td>0.39*, 0.10</td>
</tr>
<tr>
<td>Population(^c)</td>
<td>-1.09, 0.49</td>
<td>0.05, 0.11</td>
</tr>
<tr>
<td></td>
<td>0.17, 0.56</td>
<td>0.30, 0.13</td>
</tr>
<tr>
<td>(N)</td>
<td>2481, 2481</td>
<td>1564, 1564</td>
</tr>
<tr>
<td>Adjusted (R)-sq</td>
<td>0.45, 0.47</td>
<td>0.40, 0.40</td>
</tr>
</tbody>
</table>
Table 1 (Continued)

<table>
<thead>
<tr>
<th></th>
<th>Government Spending (Price Corrected)</th>
<th>Aggregate Price Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex-Western colony</td>
<td>4.02*</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>1.21</td>
<td>1.65</td>
</tr>
<tr>
<td>Ex-French colony</td>
<td>1.29</td>
<td>-1.15</td>
</tr>
<tr>
<td></td>
<td>1.42</td>
<td>-1.63</td>
</tr>
<tr>
<td>Ex-British colony</td>
<td>4.09*</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>1.26</td>
<td>-1.19</td>
</tr>
<tr>
<td>Africa</td>
<td>3.08*</td>
<td>6.37*</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>1.01</td>
</tr>
<tr>
<td>Latin America</td>
<td>-3.36*</td>
<td>2.54*</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>0.97</td>
</tr>
<tr>
<td>East Asia</td>
<td>-2.26*</td>
<td>4.33*</td>
</tr>
<tr>
<td></td>
<td>0.82</td>
<td>1.37</td>
</tr>
<tr>
<td>Years since independence</td>
<td>—</td>
<td>-0.57*</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>-0.74*</td>
</tr>
<tr>
<td>Total GDP</td>
<td>-0.06</td>
<td>-0.49*</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.11</td>
</tr>
<tr>
<td>Per capita GDP</td>
<td>-0.46*</td>
<td>5.23*</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.14</td>
</tr>
<tr>
<td>Land area</td>
<td>0.010</td>
<td>0.09*</td>
</tr>
<tr>
<td></td>
<td>0.007</td>
<td>0.02</td>
</tr>
<tr>
<td>Per capita land</td>
<td>0.24</td>
<td>1.41*</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.24</td>
</tr>
<tr>
<td>Population</td>
<td>0.45</td>
<td>-1.10*</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>-1.07</td>
</tr>
<tr>
<td>$N$</td>
<td>4521</td>
<td>4521</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.60</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Note: Results of generalized least squares. The first row in each cell represents unstandardized parameter estimates, the second standard errors. No intercepts are present because of inclusion of dummy variables for each year (estimates for dummies available on request).

†As a percentage of GDP.
* $p < 0.01$ (two-tailed test).
* Dummy variable: 0 if false, 1 if true.
* Zero for non-ex-Western colonies.
* Units as follows: hundreds of billions of dollars, inflation adjusted, for total GDP, thousands of dollars, inflation adjusted for per capita GDP, hundreds of thousands of square kilometers for land area, square kilometers per 10 people for per capita land, hundreds of millions of people for population.
On the other hand, investigation of coefficients of the various structural variables shows no clear-cut relationship between them and levels of government intervention. Total GDP is significant for the regressions involving uncorrected government spending and price levels, but its sign is positive for the first and negative for the second. GDP per capita is significant for regressions involving price-corrected government spending and aggregate price levels, but the sign for price-corrected spending is negative while that for aggregate price levels is positive.

One possible explanation for this is that detected empirical relationships between wealth and government spending in developed countries do not reflect the tendency for government output of goods and services to increase as a function of wealth. Rather, they may reflect a tendency for the prices paid by the government for goods and services to become increasingly high vis-à-vis international prices as a function of per capita wealth. This explanation is contradicted somewhat, however, by the lack of significance of the coefficients for GDP per capita in the uncorrected government spending regressions.

Total land area is significant for the regressions involving aggregate price levels, with a negative coefficient. Land per capita, on the other hand, is significant across regressions involving aggregate price levels, uncorrected government spending and spending on economic services, and its coefficient is positive throughout. Population shows no clear pattern of coefficients.

The implications of all these results for the relationship of country size versus government size and price distortion are not obvious. There are three possible measures of country size: GDP, land area, and population. With regard to distortion, higher GDP is associated with lower aggregate price levels, land area is associated with higher levels, and population is not significant. Furthermore, because aggregate price levels are an indirect measure of distortion, it is difficult to sort out influences via effects on prices of nontradeables and those via effects on government policy. With regard to levels of spending, there is again no clear pattern: higher GDP is associated with higher relative levels of government spending in one set of regressions but not in the others.

Hence, in contrast to the variable for Western colonialism, none of the structural variables displays any clear pattern across the various regressions. These results are particularly interesting because they come from the first attempt I know of to test the relationship between the structural variables and levels of state economic intervention across a large sample of both developing and developed countries.

Conclusion

The statistical analysis presented here is far from a conclusive test of the theory. Among other problems, the systematic differences between ex-Western colonies and noncolonies in numerous characteristics other than the economic structural ones includes make it difficult to rule out the possibility that the actual causation is due to some unincluded set of variables.

On the other hand, the analysis does show a striking empirical relationship between an experience with Western colonialism and a
tendency toward state economic intervention. Furthermore, there is an equally striking tendency for the level of this intervention to decline as a function of years since independence. On the other hand, there is no such striking relationship between levels of government intervention and any of the structural variables included in the regressions. This fact, combined with copious historical evidence that the state elites of ex-Western colonies did in fact internalize socialist ideology during their pro-independence struggles, provides at least reasonable support for the theory. The historical link between anti-Western nationalism and socialism is undeniable; this theory provides a plausible explanation of how it came about.

Although this does not preclude the possible existence of other, excluded ways of accounting for the pattern of outcomes revealed by the statistics, no such alternative explanation has been presented in a way that is testable across the breadth of data covered here. Meanwhile, the theory fills a large vacuum that currently exists in the explanation of the crucial historical phenomena described here, and it does so by drawing upon a wider set of hypotheses about conflict and ideology formation, which in turn are drawn from a more general coherence model as well a rational choice model of action given internalized beliefs. This in turn provides some assurance that the explanation is not an ad hoc attempt to explain a single set of phenomenon but is based on a set of deductive principles. In addition, it can form the basis for attempts to connect the phenomena of ideological formation and policy formation in ex-colonies to broader questions of ideology and policy.
Chapter 5
The Origins of Ethnic Identity and Collective Action

Ethnicity is an issue that long has occupied social scientists in general, but until recently received little attention from rational choice theorists. The reason for this is clear: the causal significance of ethnicity fits very uncomfortably with the conventional assumptions of rational choice. Since these assumptions revolve around the pursuit of economic self-interest, this implies that collective action should occur among individuals with shared economic interests. These interests in turn are reflections of individual economic characteristics, such as income, occupation and job skills. However, although ethnicity may be correlated with economic characteristics, no commonly accepted definition of ethnicity defines it as a type of economic characteristic. Hence, the causal significance of ethnicity should be nil. When collective action occurs along ethnic boundaries, this should simply reflect underlying shared economic characteristics and an explanation should be possible without any reference to ethnicity at all. Nor should there be any reason for political groups to march under the banner of shared ethnicity, since economic interest is what is really at stake.

Common definitions of ethnicity found in the social science literature relate it to primordial factors such as caste, race, ancestral homeland, language, and religion, which tend to be passed down from generation to generation. Nonetheless, ethnicity is also often seen as mutable and a product of changing environmental circumstances. Despite a long conceptual battle between “primordialists” and “circumstantialists,” much recent writing recognizes that the two definitions are complementary: some primordial commonality is a precondition for ethnicity, but the salience and level of inclusiveness of different primordial factors may vary according to circumstance. It is also generally accepted that while ethnic groups share certain cultural characteristics, culture alone is too broad a criteria by which to define ethnicity. On the other hand, ethnicity is a more general term than race, which refers to visible genetic markers and is generally defined as being immutable.

At any rate, no commonly accepted definition of ethnicity includes characteristics that appears to be directly related to economic interest, such as class, income level, production sector, or occupation. Hence,

---

1 For an overview of this debate, see Scott 1991; Stack 1988; McKay 1982.
3 Barth 1969, 10-13; Nagata 1981, 89-91. For instance, no one would consider IBM employees or heavy metal music fans to be an “ethnic group,” although they might share a common culture.
4 Rex 1986, 16-17; Young 1976, 12-13. For a more malleable view of race, see Blumer and Duster 1980, 221-23.
The Origins of Ethnic Identity and Collective Action

Ethnic collective action is frustrating to conventional rational choice theories because it seems on the face of it irrational. Nonetheless, the continuing significance of ethnicity in contemporary political behavior is impossible to dismiss. Rather than diminishing with the passage of time, ethnic variables have continued to characterize political conflict in the Third World and have experienced a resurgence both in the industrialized West and in the remnants of the former Soviet bloc. Hence, it is clear that any approach that cannot explain ethnic collective action will remain seriously deficient as a general explanatory paradigm for political behavior.

Rationalist Theories of Ethnic Collective Action

Recently, however, there have been a number of attempts to analyze ethnicity in terms that are compatible with rational choice. The resulting theories can be divided into two categories. The first is consistent with conventional rational choice assumptions about preference and belief, particularly the assumption that actors maximize their economic self-interest. The second introduces ethnic identification directly into the preference and belief structures of individuals and shows how this will influence their actions.

These theories have had two significant and complementary positive effects: Besides forcing rational choice to come to terms with ethnicity, they have also forced the study of ethnicity to come to terms with rational choice’s concern with generality and testability. Attempts to reconcile rational choice with ethnicity and, more generally, with social phenomena have constituted one of the most controversial and innovative areas of theory building in recent years. Traditionally, much of the study of ethnicity has been concerned with description or classification rather than prediction. Where predictive theories existed, they have tended to concentrate on relatively specific phenomena such as the effects of urban migration or the relationship between

---

7 An indication of the continued importance of ethnicity and other forms of group identity is the choice of “the politics of identity” as the theme for the 1992 APSA conference.
8 For discussions of the pros and cons of such attempts, see special issues of Ethnic and Racial Studies 8, 4 (October, 1985) and Rationality and Society 2, 2 (April, 1990).
9 i.e., the work of the “Chicago” school of sociologists, such as Park and Blumer as well as “assimilationists” such as Gordon.
“modernization” and the rise of nationalist movements. The rational-choice-influenced theories, on the other hand, have provided more general predictions about the factors influencing levels of ethnic collective action across different environments and time periods.

Nonetheless, I argue that these theories, despite their merits, all have inadequacies in analyzing ethnicity and ethnic collective action. Those that maintain conventional assumptions about preferences and beliefs either simply assume without proof that individuals will act collectively along ethnic lines or attribute observed instances of ethnic collective action to the existence of unoperationalized ethnic “resources” posited to facilitate individual participation. Those that introduce special ethnically based preferences and beliefs, on the other hand, cannot account for the nature and origin of such preferences and beliefs. Neither type of theory can explain the location of ethnic boundaries.

After reviewing these theories, I present a new theory, developed from the coherence-based rational choice model, that is aimed at endogenously accounting for the historical origin of large-scale ethnic group identities. It argues that individuals engaging in cooperative action with nonzero risk with members of a certain group can reduce doubt by developing a sense of altruistic identification toward those group members, which in turn will facilitate further cooperation. This theory hence not only draws upon the model of chapter 3, but also upon copious psychological writings on cooperation and the reduction of ethnic prejudice. However, the theory attempts to extend such writings to cover not only the reduction of social prejudice but also the initial formation of social identity, showing how doubt reduction can lead individuals to develop altruism toward selected others.

Rather than positing some a priori precedence for primordial factors, the theory traces the basis for identification back to early, preadult cooperative interactions in small, relatively self-sufficient communities within societies undergoing rapid structural change and increasing social mobility. While a “transitional” individual’s initial identification in such societies will be solely toward the population of those “primordial communities,” later interactions will necessitate the formation of larger groups for collective action. This will then determine how these initial altruism boundaries broadened along primordial dimensions and are transformed into larger ethnic identities. The predictions of the theory are then be used to explain the boundaries of ethnic mobilization in some of the most notable cases of “created ethnicity” during this century, including pan-Malay and pan-Igbo identities as well as the Luba-Kasai and Muhajirs.

This formulation is consistent with that taken by many theorists of ethnicity, who see ethnicity as as part of “more general processes of group formation, boundary maintenance, identity structuring or


11 For discussion of these and related issues, see Hechter 1986b.
The Origins of Ethnic Identity and Collective Action

whatever,” yet also recognize the need for “at least some minimal conceptual procedures for recognizing when ethnic identities are operative.” Given this, I hope that the theory will contribute both to a positive theory of ethnicity and to a general “theory of political cleavage” that incorporates individual motivational and cognitive factors as well as rationality.

Rational Ethnicity with Conventional Beliefs and Preferences

A number of recent authors on ethnicity have made a distinction between collective action based on special ethnic preferences and beliefs and that based on economic self-interest. For example, Young makes a distinction between “primordial” and “instrumental” ethnicity. Light between “sentimental” and “interest group” ethnicity, and Paranjpe between “subjective” and “objective” ethnicity. These distinctions correspond to the more general distinction between affective and non-affective types of social relations made prominent by the classical social theorists such as Tonnies, Durkheim, and Weber. They also correspond somewhat to the distinction often made between “primordial” and “circumstantial” ethnicity, though associative ethnicity is somewhat narrower than circumstantial, excluding affective components, whether they be based in the distant past or the present.

Although some theorists may condescendingly call the latter “pseudoethnic,” others have taken it to a position of preeminence in explaining ethnic collective action. They imply that there is no need to resort to altruism or expressive benefits as explanatory variables when economic self-interest can generate more persuasive arguments.

A number of prominent theorists have recently incorporated ethnicity into theories that are compatible with conventional rational choice assumptions.

The Cultural Division of Labor

Hechter’s theory of the “cultural division of labor” revolves around the proposition that ethnic collective action will be most intense when ethnic boundaries align with particular economic boundaries. In the original version of his theory, he posits that pronounced, systematic differences in wealth between different ethnic groups will create the

---

12 Mason 1986, 11. See also Hechter 1986a, 265.
13 Rogowski 1985b, 384.
14 Young 1983b, 660.
15 Light 1981, 55.
16 Paranjpe 1986, 2.
17 In particular, the two types of ethnicity correspond very closely with Weber’s distinction between “communal” versus “associative” relations. See Weber 1978, vol. 1, 48.
18 Scott hence divides circumstantial ethnicity into two subcategories, one based on “structural conditions” and one on “rational, strategic selection.” See Scott 1991, 148.
20 Hechter 1987b, 11.
perception among members of the low-income group that their life chances are limited by their ethnicity. As a result, they will develop a shared economic interest in mobilizing politically to overturn these wealth disparities. In a later version of the theory, Hechter adds the proposition that shared interests will be generated when members of a particular ethnic group cluster within a particular occupation, whether or not this occupation niche is associated with a particular income level.

The undeniable asset of the theory is that it can be used to predict variations in the levels of collective political action within a particular specified ethnic group over time as well as between different ethnic groups. On the other hand, it does not attempt to predict the location of ethnic boundaries, which are taken as exogenous. Its basic logic predicts that shared income or occupational boundaries among members of an ethnic group will create common interests, which in turn will promote collective action. Without further elaboration, this would imply that individuals within such economic boundaries will have a tendency to act collectively, regardless of whether they share any ethnic characteristics.

Hechter clearly does believe that such characteristics are significant in their own right. To justify this, he suggests at certain points that a common culture lowers the costs of collective action among individuals, stating that:

> cultural differences between individuals generally impose barriers to communication between them. Language is the most obvious of these cultural difference, but it is by no means the only one. The interpretation of meanings across boundaries is invariably problematic.

Although this is certainly a reasonable statement, he makes no systematic attempt to operationalize cultural variables or specify their precise impact on individual incentives and collective outcomes. Although Hechter undertakes a number of clearly specified analyses of individual incentives in collective action in a later book, they are

---

21 First presented in Hechter 1971, 21-45 and elaborated in Hechter 1975. These ideas have antecedents in ideas put forth by Gellner in “Nationalism” in Althought and Change, chap. 7; see also his Nations and Nationalism exp. chapter 6. Similar ideas are presented in Shibutani and Kwan 1965 and Nairn 1977. It also is analogous to arguments that ethnic conflict will be created by “uneven modernization” among ethnic groups. See Melson and Wolpe 1970, 1115-17. See Bates 1974, 464-65, which reappeared in modified form as Bates 1983b. Horowitz makes the somewhat different argument that conflict in “ranked” systems will be less frequent but more intense than in unranked systems. See Horowitz 1985, 30-35.

22 Hechter 1978. The latter type of cleavage is called “segmental,” while the former is called “hierarchical.” One precursor Hechter mentions for this idea is Barth 1969, 16-17, who posited that ethnic boundaries will remain most distinct if groups occupy distinct niches in their environment.

23 See Olzak 1983, 360. Note below, however, that Olzak’s own theory shares an analogous tendency.


not directly related to ethnic collective action. In a later essay, he ties together his analyses of ethnicity and group solidarity, but ethnic group boundaries remain exogenous.26

Because ethnic boundaries remain exogenous, they must be specified in an ad hoc manner. Hence, the predictions of the theory are clearest for the case in which these prespecified ethnic boundaries and economic boundaries are closely aligned. In other cases, it does not attempt to predict what characteristics will determine the boundaries of collective action or what level of collective action will occur, hence the theory cannot be easily be tested against theories of collective action that ignore ethnicity.27

Modernization and Competition

The “competition model,” associated with authors such as Olzak and Nagel, links ethnic collective action to “niche overlap,” which is created by the breakdown in ethnic occupational barriers associated with the increased mobility and economic rationalization resulting from modernization. Ethnic groups move into job markets where they were formerly absent, triggering competition with other ethnic groups that already occupy these jobs. Members of the entrenched ethnic group develop a common interest in excluding the upwardly mobile ethnic group from their niche, while the members of the upwardly mobile group develop a common interest in ensuring their access.28

Hence, the theory predicts that ethnic collective action will occur when ethnicity aligns with the economic boundary between workers entering an occupational niche and those who are entrenched within it. The theory is often viewed as contradictory to Hechter’s, though this is not necessarily the case; both the absolute degree of ethnic segmentation or hierarchy and the rate of its change could be factors in promoting collective action.29 In fact, the competition theory faces an analogous dilemma to that faced by Hechter’s theory; it must show

26 Hechter 1987a, 424-25.
27 At one point, Hechter posits that collective action will occur along class lines rather than ethnic lines if class is more highly correlated with income or occupation. However, this leads to more questions: what determines the boundaries of class, if not income or occupation? Does class also possess cultural resources? If not, what accounts for its independent causal significance? What happens if neither class nor ethnicity is aligned with either economic boundary? See “Group Formation,” 307-8.
28 The initial version of the theory was put forward by Hannan in “The Dynamics of Ethnic Boundaries in Modern States,” in Meyer and Hannan 1979 and developed in Nagel and Olzak 1982 and several subsequent articles. The authors of the theory, like Hechter, attribute influences from Barth 1969, 20 cited in Olzak 1990, 27. They also have early precursors in the ideas of Park on the first two stages of the “race relations cycle.” See Park 1950, 149-51.
29 In fact, Melson and Wolpe as well as Nielsen, incorporate analogues of both hypotheses in their propositions about modernization and ethnicity. See propositions 2 and 3 in Melson and Wolpe 1970, 1115-18 and propositions 6 and 9 in Nielsen 1985, 142-44.
why competition for jobs would not tend to produce conflict between newcomers and incumbents regardless of their ethnicity.\textsuperscript{30}

Like Hechter, Olzak adds some justifications for why ethnicity might have independent causal significance. In particular, she focuses on the organizational resource provided by ethnicity in the form of “interconnected networks, institutions and information systems” built up over time through stable interactions.\textsuperscript{31} While this provides another plausible reason why ethnic collective action might occur, she does not operationalize organizational variables or specify precisely how they affect incentives. Hence, as with Hechter’s theory, the boundaries of ethnic groups must be specified in an ad hoc manner. Predictions about the boundaries and levels of collective action are possible primarily when exogenously determined ethnic boundaries and traditional occupational niche boundaries align, and the latter are breaking down.

\textit{Split Labor Markets}

Bonacich’s theory of ethnic antagonism attributes conflict to a market in which laborers of different ethnicities receive different wages for the same work.\textsuperscript{32} This initial wage gap can be caused by many factors: uneven development, imperialism, migration, and the tendency of certain minorities to play “middleman” economic roles.\textsuperscript{33} Because of their common interest in keeping members of the lower-waged ethnicity from receiving comparable wages, the members of the higher-wage ethnicity organize to protect their advantages through exclusion and confrontation. This division among the laboring ethnic groups, however, ultimately redounds to the advantage of capitalists (from either ethnic group), who can use it to play the two groups off against one another in order to minimize wages for both.

This theory has clear similarities to both the competition and cultural division of labor models,\textsuperscript{34} but, unlike Hechter, Olzak and Nagel, Bonacich states clearly that ethnicity does not have independent causal significance. Instead, it can form the basis for collective action only because it is frequently aligned with differences in the price for labor.\textsuperscript{35} Hence, Bonacich addresses the question of ethnic versus economic boundaries by giving precedence to the latter. This implies that when ethnic boundaries and wage boundaries do not coincide, ethnic collective action will not occur.

However, this seems inaccurate given the frequency with which collective action and conflict occur along ethnic lines despite the

\textsuperscript{30} See related criticism in Hechter’s review of Olzak and Nagel 1986; Hechter 1988a, 108.

\textsuperscript{31} Olzak 1985, 80.

\textsuperscript{32} Bonacich 1972; 1979.

\textsuperscript{33} For more on the last factor, see also Bonacich 1973.

\textsuperscript{34} It can be seen as conjunction of the two in that both income differences between ethnic groups and niche overlap need to be present in order for collective action to occur.

absence of a clear-cut alignment between ethnicity and income.\textsuperscript{36} Furthermore, even when some alignment exists, it is usually not perfect, and the theory cannot explain why action by high-wage laborers should take place under the banner of ethnic exclusion rather than job security for incumbent labor, regardless of ethnicity.\textsuperscript{37} Finally, even in an ethnically split labor market, it is not clear why higher-wage and lower-wage laborers would not unite and organize if they realize that their conflict is mutually hurting them vis-à-vis the capitalists rather than battle along ethnic lines.\textsuperscript{38} Hence, while nullifying the causal significance of ethnicity may remove the problem of explaining its role, it makes much of ethnic collective action seem anomalous.

**Factor Monopolization and Discrimination**

Banton’s theory of rational choice ethnicity contains the key proposition that ethnic groups monopolizing certain goods can increase their aggregate wealth by centralizing control over trade with other groups, hence gaining monopoly profits.\textsuperscript{39} This in turn requires barriers to individual-level trading since individual members can increase their income by selling the good for less than the monopoly price. These barriers can take the form of group rules that mandate discrimination in interactions with nongroup members; an example would be rules prohibiting white employers (who monopolize the means of production) from paying black employees the same amount as white employees.\textsuperscript{40} However, this theory provides no clear explanation of why discrimination should occur along ethnic lines. As long as individuals are free to act collectively along lines of their own choosing, the logic of the theory suggests that individuals within particular production or technical skill sectors should organize for collusion, whether or not they are all members of a single ethnic group. Furthermore, it would be in the interest of monopolists to discriminate against other members of their own ethnic group as long as they belong to a different sector. Banton provides no clear justification for why ethnicity should be the chosen boundary for discrimination, nor does he provide a theory for predicting the location of ethnic boundaries.

\textsuperscript{36} e.g., Sikhs in the Punjab, Tamils in Sri Lanka, Mohajirs in Pakistan, Basques and Catalans in Spain, Oromos in Ethiopia, Ovimbundus in Angola, Karen and Mon etc. in Burma; multiple groups in the ex-Soviet Union . . .

\textsuperscript{37} Such actions may in fact have the result of lowering the wages for those few members of the nonfavored ethnicity who are in the high-wage group, as happened in South Africa after the passage of the Color Bar Act.

\textsuperscript{38} Bonacich attributes this to both the possibility of false consciousness and coercive barriers to the organization of lower-wage workers, though she does not discuss this in detail. See “The Past, Present, and Future,” 34-35. The former is ad hoc and begs the question: what accounts for false consciousness? The latter seem hard to maintain in for any length of time if both sets of laborers work in the same market, unless high-wage laborers want the barriers to remain.

\textsuperscript{39} Banton 1983, chap. 6.

\textsuperscript{40} Banton 1983, 119-20.
Pillarization and Nationalism

Rogowski has constructed a general theory of government legitimacy, which has imbedded in it certain propositions about ethnic collective action.\(^\text{41}\) Although the connections between legitimacy and action are not always clear in his original formulation of the theory, it is strongly suggested that tendencies toward separatist collective action will be strongest among “pillars,” i.e., groups with the requisite skills to form a self-sufficient independent state.\(^\text{42}\) In recent writings, Rogowski makes the connection more explicit.\(^\text{43}\)

In addition, Rogowski implies very strongly that nationalist collective action will take place along ethnic lines. In fact, all the illustrative discussions of factional behavior in his writings concern ethnic groups. As a justification for this, he posits that factions will have greater internal discipline if members are characterized by stigmata, i.e., human characteristics that are easily identifiable and changeable only at high cost. These stigmata make it more difficult for members of a faction to engage in noncooperative behavior without being detected, and hence they facilitate the application of intragroup sanctions.\(^\text{44}\) However, he makes no attempt to operationalize the extent of shared stigmatism in a way that is easily measurable. Furthermore, the ability to quickly distinguish group members from nonmembers in order to apply sanctions seems on the face of it a rather minor factor in a large group’s ability to act collectively, no matter what specific type of action is being considered. If it were so important, we would expect ethnic groups to frequently form around characteristics such as height and weight and rarely if ever around such characteristics as religion or region of origin. Rogowski himself provides no concrete examples of how common stigmata promote collective action; his only illustration involving stigmata describes how the migration of individuals into new occupational niches may have caused the decay of stigmata, but not the effects of that decay.\(^\text{45}\)

The Politicization of Ethnicity

One final set of theories in this category that ought to be discussed emphasizes how state policies, especially those that divide populations up into administrative units based upon particular characteristics and/...
or those that favor groups with one set of characteristics over others, can be a force promoting ethnic collective action. This has been most definitively stated by Nagel\(^{46}\) and Brass\(^{47}\), but has also has occupied much of the writing of other prominent scholars such as Young\(^{48}\) and Horowitz.\(^{49}\) The basic logic shared by these authors is that states, by making particular individual characteristics a basis for the allocation of economic benefits, provide an incentive for individuals sharing those characteristics to act collectively.

The main difference between this and the other theories being discussed is that the categories that divide the population along economic lines may sometimes correspond directly to primordial characteristics such as race or language. In that case, no additional explanation is required to explain why collective action will occur along ethnic boundaries or to determine the location these boundaries since common economic interests will coincide exactly with the particular ethnic boundaries specified by these policies.

Nonetheless, these theories leave the state as an exogenous actor, and it cannot explain in conventional rational choice terms why it would engage in policies that divide individuals according to primordial characteristics rather than others. Their logic simply implies that states have the power to influence the boundaries of collective action, whether the boundaries it promotes are ethnic or not. Of course, it may be true that populations are often most receptive to mobilization along primordial lines than others, but the theories provide no clear way of explaining this phenomenon.

**The Problem of Ethnic Resources**

These theories have unquestionable and important assets: each can be used to predict variations in levels of collective action along particular ethnic boundaries within a broad range of socioeconomic and political environments. Each is more or less consistent with conventional rational choice assumptions about individuals, and hence they retain some of the deductive qualities of the conventional model. Because of this, they have brought a new level of generality and rigor to the analysis of ethnic collective action.

However, one notable characteristic shared among them is the limited amount of attention they pay to problematizing ethnicity. Their common basic structural assumption is that if an ethnic group’s boundary aligns with some economic boundary (a common position in the division of labor, upward mobility, possible self-sufficiency), individuals in the group will have common interests that will provide a basis for collective action. However, these assumptions by themselves would lead to the conclusion that collective action will tend to occur

---

46 Nagel 1986.
47 Brass 1991, intro., chap. 7
48 See his chapters in both the Olzak and Nagel and Brass volumes as well as Brass 1982.
49 See Horowitz 1985, especially chapters 4 and 5 as well as Horowitz 1975.
occur along these economic boundaries even in the absence of ethnic divisions or despite ethnic divisions that crosscut them.\(^{50}\)

Because of this, those theories that posit that ethnicity is independently significant then invoke some notion of ethnicity-related “resources” in the form of culture or organization to explain why collective action will occur along ethnic boundaries. But these resources are not theorized systematically. No methodology is provided for determining the levels of various resource variables or predicting their effects on individual incentives. As Rex notes,

The curious feature of such theories is that they do not seek to explain the existence of groups, quasi-groups or the ethnic resource. They simply take it for granted.\(^{51}\)

Hence, the notion of ethnic resources cannot be used to test these theories against those that ignore ethnicity and concentrate solely on collective action by individuals with shared economic interests.

Furthermore, these theories do not attempt to predict the salient boundaries of ethnicity for particular types of political action or how they change over time. Instead, the boundaries are left exogenous. When discussing actual cases, the authors take existing distinct groups (e.g., Scots, American blacks, or Tamils) as their units of analysis. There is a certain circularity in this approach, since the mere presence of these groups as recognizable entities indicates some prior existence of ethnic mobilization.

It is conceivable that the problem of boundaries could be resolved by a more broad and precise specification of the ethnic resource and more detailed reasoning about how it affects individual-level incentives, but there are inherent limits in such an approach. The ways in which culture, language, physical appearance, existing organizations, and so forth can conceivably promote cooperation are so numerous and varied that any attempt to pack them into a single concept would render it unmeasurable except tautologically.\(^{52}\) On the other hand, any attempt to single out a particular variable, such as language, frequency of past interactions, or physical similarity, risks reductionism to the point of inaccuracy since for any one variable there will be numerous examples in which it cannot be used to explain the boundaries of collective action.

Finally, no conception of ethnicity based on resources and conventional economic self-interest can be consistent with the apparent motivations that drive individuals participating in many types of ethnic collective action. According to Epstein, the underlying issue is “the powerful emotional charge that appears to surround or to underlie so much of the ethnic behavior, and it is this affective dimension of the problem that seems to me lacking in so many recent attempts to handle

\(^{50}\) For related criticisms, see Connor 1984.

\(^{51}\) Rex 1986, 29.

\(^{52}\) The same holds for attempts to develop a concept of “social capital.” See Coleman 1990, chap. 12.
it. As Bell notes in his oft-quoted statement, ethnicity is a powerful determinant of collective action because it can “combine interest with an affective tie.” As Rothschild puts it, ethnicity’s great appeal “lies precisely in its ability to combine emotional sustenance with calculated strategy.” Or, as Glazer puts it, ethnicity contains both a rational component and “a less rational, an irrational [sic] appeal, that seems to connect better with powerful emotions.” Douglass argues that, if we are to progress beyond our current rather mechanistic and uninspired models we must be prepared to address the elements of affect and altruism germane to particular nationalistic traditions, and in their own terms.

Finally, as Nagata puts, it,

Ethnic identity then is a unique blend of affective, expressive and basic ties, sentiments and loyalties with (sometimes bluntly) instrumental, calculated, political interests, and the latter are explained and given meaning by the former. The two ends of the Gemeinschaft-Gesellschaft continuum here come full circle.

Ethnically based motivations are not necessarily irrational, but they imply that individuals have affective preferences that go beyond the conventional assumptions of economic self-interest. Without taking such preferences into account, rational choice theories can never account for the cases of self-sacrificing behavior that are hardly uncommon in cases of ethnic collective action. Some of the authors discussed in the previous section do hint at the possible existence of motivations that go beyond individual self-interest. For instance, Olzak writes that ethnic enclaves “reinforce identification with cultural symbols in an ethnic community, such as language, customs, religion and other practices.” Hechter and Levi in turn write that cultural factors can provide nonmaterial benefits, such as “opportunities to identify oneself as a member of a cherished ethnic group.” However, none of these points is developed systematically and incorporated into the theories they present.

Ethnicity as Special Preferences and Beliefs
Rather than attempting to assimilate theories of ethnicity into the melting pot of conventional rational choice, other theorists have decided

53 Epstein 1978, xi.
54 Bell 1975, 169.
56 Glazer 1981, 79.
57 Douglass 1988, 204.
58 Nagata 1981, 112.
60 The Dynamics of Ethnic Competition, 40.
61 Hechter and Levi 1979, 269-70.
Choosing an Identity

to explicitly view it as an expression of special, unconventional types of preference and belief:

**Discriminatory Preferences**

Some of earliest work in this field was done by Becker, who assumed a “taste for discrimination” among individuals, which is manifested as a “discrimination coefficient” which increases costs to transactions with individuals of a disliked group.62 Becker then examined effects relating to discrimination among a variety of actors in a variety of market environments, analyzing such things as the effect of discrimination on group aggregate incomes, the effect of changes in relative population on relative incomes, the effect of labor and output market concentration on levels of discrimination, and the causes of workplace and residential segregation.

Johnson, investigating the policies of Third World leaders, posited a “taste for nationalism” linked to having members of one’s own nationality hold jobs or own property.63 Given such preferences, he argued, it can be personally rational and even collectively welfare enhancing for leaders to carry out policies of discrimination and confiscation, even when they lower aggregate domestic wealth.

Theory building in this vein has been continued by Sowell, who notes that, given tastes for discrimination, actual discrimination against undesired groups will be most pronounced where wealth-maximization incentives dictating against discrimination are low (e.g., in nonprofit firms).64 In addition, Klitgaard notes that the known presence of a percentage of individuals with discriminatory tastes may cause even individuals without such tastes to discriminate, because they can sanction those who violate group norms mandating discrimination.65 In addition, Schelling shows how even slightly discriminatory preferences can lead to highly segregated patterns of residence.66

Although it is quite plausible that individuals will have discriminatory preferences, the shortcoming of all these theories is that none provides an assertive set of assumptions about the nature or origin of such preferences. None can predict the extent to which any individual will have discriminatory preferences or whom those preferences will be

---


64 Sowell 1975, 165-67.

65 Klitgaard 1976, 9-11. Klitgaard somewhat confusingly calls those with discriminatory tastes “irrational racists,” presumably because he feels that nonwealth-maximizing preferences should not be considered rational. On the other hand, however, he considers acting according to social norms a kind of rational behavior (“rationality-2”).

66 Schelling 1978, chap. 4.
directed against. Without a way to determine this, the theories cannot be used to predict actual behavior.\footnote{One alternative might be to use some sort of “revealed preference” methodology (as with theories of consumer choice) to determine levels of discrimination, but a problem with this that individuals, unlike inanimate goods, cannot be classified as uniform commodities. Since no two individuals are exactly alike, it is not straightforward to translate discrimination between individuals into conclusions about discriminatory preferences. At any rate, none of the theories includes such a methodology.}

**Altruistic Preferences**

Given the relatively large amount of work on modeling altruism in theories of collective action,\footnote{See citations in section on multidimensional utility function in chapter 2.} it is surprising that there is little work in doing so for theories of ethnicity. In fact, the only prominent work that uses such an approach does so only implicitly. Rabushka and Shepsle were the authors of one of the earliest efforts in the field of rational choice ethnicity, and their work remains one of the most elaborate efforts at theory development.\footnote{Rabushka and Shepsle 1972.} Ethnic boundaries in their theory are exogenous, and most of their theoretical discussion revolves around the abstract case of a biethnic society. Individuals are assumed to have a one-dimensional utility function based on the closeness of political outcomes to their ethnic group’s (unspecified) ideal point. In other words, individuals will care only about solely their group’s “preferences” and not independently about their own welfare. This in effect means that they are completely altruistic toward their group and its members. In addition, in “intensely divided” societies, individuals are risk accepting with regard to these outcomes. This in turn causes them to prefer lotteries (conflict between groups) over sure outcomes (compromise), given equal expected values. Political entrepreneurs, realizing this, will “bid up” their appeals to ethnic hostility, knowing that this will garner support.\footnote{Rabushka and Shepsle 1972, chap. 3.}

Although the assumptions are more assertive than those in theories of discriminatory preferences, the assumption of complete altruism toward the group is implausible and produces inaccurate predictions since it implies that within-group conflict would never occur. These extraordinary utility functions in effect eliminate the collective action problem. Groups can be treated as unitary entities since all individuals within them will have identical preferences over outcomes. The model becomes, in effect, an expected-utility war model.\footnote{Bueno de Mesquita 1981.} Furthermore, Rabushka and Shepsle provide no clear reason why risk-seeking should follow from intense divisiveness.\footnote{see Rabushka and Shepsle 1972, 49-52.} Given intense dislike for another ethnic group, it is quite plausible than it would be more painful to live under its dominance that it would be pleasurable to dominate it. Conversely, in an undivided society in which there is no
disagreement between groups over outcomes, it is clear that individuals
would not be risk averse but rather indifferent between risk and safety
since they would not mind at all if the other group dominates.

Finally, the assumptions cannot be used to predict the boundaries
of ethnic groups nor the kind of preferences that will be associated
with different groups. Hence, given the effective view of groups as de facto
unitary entities and the additional assumption of risk seeking, the only
prediction that can be made is that of the inevitability of conflict.

Cultural Beliefs
Another set of theories incorporates ethnicity into beliefs rather than
preferences, positing that cultural beliefs with a particular ethnic
group can facilitate cooperation along ethnic lines. These theories are
confined to single cases, and since they do not provide methodologies
to account for the origin of these beliefs, they cannot form the basis
for general theories of ethnic collective action. However, they can shed
light on how particular types of beliefs can alter individual incentives
within a particular group.

Landa argues that, in the absence of enforceable contract
law, common cultural beliefs within ethnic groups can provide
coordinating mechanisms that ensure cooperation and prevent breach
of contract. Hence, trading networks in countries without established
legal institutions will be dominated by ethnically homogenous middlemen.73 Greif makes a similar argument, positing that shared
collectivist beliefs among Magribi traders in the eleventh century made
them believe that anyone who breached contract against a single
Magribi would cheat them all. Hence, traders would invest in setting
up information networks and no trader would have an incentive to
breach a contract since no other trader would be willing to trade
with him subsequently. On the other hand, where individualist beliefs
existed, traders would ignore the previous behavior of other traders and
cooperation would not be possible without centralized enforcement.
In sum, “collectivist cultural beliefs...can sustain cooperation in
situations where it cannot be sustained by individualist cultural
beliefs.”74

Explanation of Ethnic Preferences and Beliefs
Although all of these unconventional theories manage to demonstrate
ways in which ethnicity can be incorporated into individual preferences
and beliefs, they cannot account for the origin of these preferences and
beliefs or the boundaries within which they will be shared. Hence, the
predictive power of such theories is limited unless a methodology is
provided that can explain such origins in an endogenous manner.

Fixed Primordial Altruism
One possible way to endogenize ethnic preferences is to assume a fixed
amount of altruism toward others based on some fixed set of primordial
characteristics. One version of this would be to base altruism on some

73 Landa 1978; see also Landa 1995.
The Origins of Ethnic Identity and Collective Action

notion of genetic kinship, as might be suggested in the writings by Van den Berghe that link ethnic collective action to sociobiology.\footnote{Van den Berghe 1978; 1979. For criticisms and defenses of this approach, see Reynolds 1980, and a debate that followed in that journal as well as Van den Berghe 1986.}

The most natural way of operationalizing this in rational choice terms would be to assume that the weight that an individual places on the welfare of another individual (including himself or herself) is proportional to the percentage of genes shared with that individual. Such assumptions could be used to explain certain types of action that cannot be explain by assumptions of narrow self-interest. Nonetheless, the usefulness of such an approach would be quite limited. For any kind of cooperative action that extends beyond a small extended family, the percentage of genetic material shared between an individual and those who would benefit from his or her action multiplied by the amount by which the individual’s actions could benefit them would be much too small to have anything but a minor effect on an individual’s incentives. For actions on a national level, the effect would be miniscule.

Furthermore, kinship altruism cannot account for the fact that ethnic groups are often based upon primordial factors other than kinship, such as language, religion and geographical region. While it may be true that “markers” are often correlated with kinship, this does not explain why they, rather than kinship itself, should be the determinants of group boundaries. Although it may be argued that these markers serve as the basis for ethnicity because kinship itself is difficult to determine,\footnote{“Race and Ethnicity,” 406.} it is not clear why this should be so except in the most spontaneous types of interactions. Furthermore, it does not explain then why ethnicity often includes such phenomona as religious conversion and assimilation, or why the boundaries of self-identification may shift over time.

A more nuanced model of ethnic preferences might focus on altruism based upon several different types of primordial factors rather than just one. For instance, a theory of altruism might be adapted from Isaacs’s framework, which examines body, name, language, history and origins, and religion, each of which are posited to constitute elements of “basic group identity.”\footnote{See Isaacs 1975a and Isaacs 1975b.} The main unresolved issue with such an approach is explaining why different types of primordial factors are important under certain types of conditions, while other primordial factors (e.g., height and weight) never form the basis for ethnicity. This in turn requires a theory that acknowledges the significance of these factors but also explains how their influence varies across cases and over time.

Elite-inspired Ethnic Consciousness

Another body of literature that might be used to endogenize ethnic group-specific preferences and beliefs is that which emphasizes the role of elites in promoting, for various reasons, various types of ethnic
Choosing an Identity

consciousness among nonelite members of their groups.\footnote{For rational choice-type approaches, see Rogowski 1985a, 99-102 and Meadwell 1989a, 318-23. There is a much larger literature outside of rational choice which centers around the elite role in promoting ethnic ideologies. See Smith 1991a and Smith 1981a as well as Brass 1991, 25-36 and Royce 1982, 105-7.} Such writings provide descriptions of the conditions under which elites will have an incentive to promote group consciousness (which can take the form of within-group altruism and beliefs in the existence of common goals) among nonelites with which they share primordial characteristics. However, as noted in chapter 3, such a view of elites as idea entrepreneurs cannot be used to predict nonelite beliefs without a theory that can explain when particular elites will be successful in having the masses adopt their views. Simply assuming that elites will always be successful in fostering ethnic identification requires viewing the recipients of such efforts as implausibly malleable, and an assertive set of alternative assumptions has not been offered within this literature. Nor does it explain why groups that share primordial characteristics will be more susceptible to “consciousness-raising” than groups who do not. In its current state, such an approach cannot form the basis of an assertive theory.

Hence, a review of existing rational choice theories of ethnic collective actions shows that conventional theories cannot adequately explain why collective action should along ethnic lines, while theories incorporating unconventional preferences and beliefs cannot explain their origin. Because of this, each type of theory must take ethnicity itself as exogenously determined, and it cannot explain the formation of and changes in ethnic group boundaries. What is needed, then, is a way of accounting endogenously for ethnic preferences and beliefs in a way that is compatible with the assumption of rational action.

A Coherence-Rational Choice Theory of Ethnic Group Formation

The theory of ethnicity presented here attempts to address some of these shortcomings of existing theories by accounting endogenously for ethnically-based preferences, but in a way that is both assertive and can predict the levels and boundaries of collective action. It is influenced by existing theories of rational ethnicity, but unlike them attempts to account for ethnic preferences and their effects on collective behavior.

Ethnic preferences are posited to stem from group altruism, the incorporation of a particular set of individuals’ welfare into another’s utility function. Altruism is the simplest and the most general way of modeling ethnic group preferences within rational choice assumptions. It allows one construct to be used to specify both the boundary and intensity of ethnic group preferences over all cases, without having to focus on any single primordial factor. Furthermore, unlike many alternatives definitions of identity, it has clear consequences for action.\footnote{Hechter 1988b, 18.} The definition also is consistent with the “fellow-feeling” that has commonly been associated with group identity in the psychological
literature and is also consistent with recent practice in anthropology of basing ethnicity on subjective boundaries rather than on differences in particular cultural traits or practices. Finally, it has a number of characteristics that allow it to be used for relatively nuanced analysis: it allows individuals to identify with more than one group (including identity “layers” of varying intensities), it allows for the possibility that individuals may identify with groups of which they are not members, and it allows for changes in identity within individuals over time.

The theory posits that group altruism formed within individuals as a result of past cooperation. In societies that have not experienced structural changes such as economic commercialization or the consolidation of large-scale political units, such identities will be directed within the boundaries of small, stably populated, and relatively self-sufficient communities. This, however, will change at points in history when patterns of cooperation have been enlarged dramatically due to the aforementioned structural changes. For that generation of individuals whose childhood interactions are restricted to small communities but whose adult experiences are in larger arenas, new types of affiliations will be based upon specific interaction between their existing community-level identities and the need to organize into larger groups for effective economic and political cooperation in the urban metropole. Eventually, these new affiliations will be internalized into altruism patterns and become the basis for large-scale ethnic identities.

The logic of this theory is used to generate a number of hypotheses concerning the timing and boundaries of large-scale ethnic group formation. These hypotheses in turn are tested upon a large number of empirical cases of ethnic group formation that have taken place during this century.

Cooperation and Altruism in Primordial Communities
The historical origin of large-scale ethnic groups can be traced to societies in which the majority of the population still resides in small, stably populated, and relatively self-sufficient communities but where structural changes have taken place to the extent that individuals will eventually engage in patterns of economic and political cooperation far beyond the borders of their communities of origin. This “transitional period” of structural change been a major focus for analysis by both nineteenth century social theorists and by theorists in the

---

80 Most notably Tajfel’s theory of social identity, see Tajfel 1981. See also DeVos 1975, for a discussion specific to ethnic groups.
81 See famous passage from Barth 1969, 10-16.
82 See for instance the discussion in Horowitz 1975, 119.
83 e.g., Mallie’s “white negro.”
84 In this way, it is a “Lamarckian” view of ethnicity rather than a “Mendelian” one. See the introduction to Linnekin and Poyer 1990.
Despite disagreements over the accuracy of the specific predictions made by such theories, there is no question that the transition from small communities to larger arenas of interaction is one of the most significant effects of long-term structural change.

It can reasonably be generalized that throughout most of history patterns of cooperation have revolved around a relatively small community of individuals, whether they be members of an extended household or of a village whose membership remained relatively stable over time. Furthermore, because there was relatively little division of labor, it can be posited that each cooperative activity involved the bulk of the members of these communities. Such communities correspond very closely to a Gemeinschaft and, for want of a better term in English, will be referred to as “primordial communities.”

Given such a pattern of cooperation, it can be inferred that an individual’s initial altruism will be directed toward members of his or her community. The reasoning behind this inference follows, at the micro-level, from the coherence model. In a primordial community, collective action among community members will be crucial to the provision of goods needed for their welfare, such as defense against outsiders, provision for preparing crops, protection against natural disasters, and building of shelters. In such collective action, however, there is often a temptation for individuals to free-ride, i.e., not participate in collective labor and attempt to nonetheless enjoy the benefits that come from the cooperation of other community members. However, the close-knit nature of the community means such temptations to free-ride will be more than offset by the ease in which their behavior can be monitored and punishments applied. Hence, rational individuals will in such communities will for the most part behave cooperatively. Nonetheless since monitoring cannot be perfect, there will always be a nonzero probability that free-riding behavior might have paid off (i.e., that shirking might not have been detected). Hence, decisions to cooperate will generate a certain amount of doubt, because of the nonzero probability that cooperative behavior was inconsistent with the individuals’ own interests.

If this occurs, the development of altruistic preferences toward community members can serve as a kind of doubt-reduction device because it allows individuals to reduce possible doubts that they may have over their cooperative actions. If individuals intrinsically value the contribution their efforts make to others in the community, cooperative behavior may be rational even in cases in which actions did not bring them any marginal personal benefit. Hence, from the discussion in chapter 3 on theorem 3 and altruism, it can be inferred that an individual who repeatedly cooperates under such circumstances will tend to develop a sense of positive economic altruism toward members of the community.

This type of inference is also consistent with the “contact hypothesis” put forward by Allport and others in their research on

---

86 Lerner 1958; Riggs 1964; Pye 1962, chap. 2; Deutsch 1963.
race relations, which links altruism between individuals directly to past patterns of cooperation.\footnote{See citations in chapter 3 in the discussion of applications of the coherence model.} Furthermore, it accords with experiments that show that an individual’s liking for another individual can be determined by whether their own past actions toward the other are seen as cooperative or conflictual.\footnote{See Davis and Jones 1960.} Finally, the inference is in accord with the general ideas of the classic social theorists, who believed that strong affective ties between individuals were a distinctive characteristic of “traditional,” i.e., subsistence, communities.\footnote{The classic treatment can be found in Tonnies [1887] 1957, 42-44.}

Of course, this theoretical excursion does provide a simplified and somewhat idyllic picture of life in primordial communities. There is clearly conflict as well as cooperation in such communities, and in reality an individual’s level of economic altruism will vary toward different members of the community and may even be negative toward some. However, it is also true that a typical individual’s economic actions will in aggregate have a positive net effect on the economic welfare of other community members. This can be shown by comparing the expected economic welfare of individuals in isolation to their expected economic welfare given interactions within a community; it is clear that the latter will generally be higher. Furthermore, the pressure of the environment in such communities is usually high enough that any failure to cooperate could quite possibly lead to mutual starvation. This in turn increases the shared interest of group members in the maintenance of cooperative outcomes, even if it does not totally eliminate the temptation to freeride. If this is true, then positive altruism should on the whole outweigh negative altruism.

To simplify quite a bit, we can model the utility function of an individual in a primordial community as $u_i + a_c u_c$, where $u_i$ represents the individual’s economic welfare, $u_c$ represents the aggregate economic welfare of the individual’s community, and $a_c$ represents the weight of altruism toward the community, with $a_c > 0$.\footnote{It might be more realistic to add additional variables to represent an individual’s somewhat higher levels of altruism toward his or her immediate family as compared to extended family members and nonkin, but they will tend to generate complications without any effect on the logic of the theory.} However, such altruism cannot reasonably be called “ethnic” since the population of individuals toward whom an individual will feel altruism is much too small to be considered an ethnic group. Furthermore, while members of the primordial community may share primordial characteristics such as language, religion, race, and so forth, individuals will at this point not need to demarcate the boundaries of their altruism based on these characteristics. Their sense of altruism will be attached to the concrete
Choosing an Identity

group of individuals attached to their primordial community, not to any abstractly defined “imagined community.”

The Expansion of Group Boundaries in the Metropole

As noted, the focus of the theory is on societies where changing structural conditions provide individuals with incentives to migrate from their primordial communities into larger spheres of interaction. The major structural changes that create such incentives are economic commercialization and the consolidation of large-scale political units.

Commercialization is defined here as the tendency for units of economic production to produce a narrow range goods on a large scale and for members of those units to trade these goods for other desired goods. Political consolidation is defined as the construction of effective state authority systems that span very large numbers of primordial communities as well as their associated systems of wealth allocation and redistribution. Obviously, these two phenomena are not completely independent. However, the two types of change can clearly move forward at different rates.

The causes behind these two structural changes can be multifaceted; the most obvious cause is endogenous technological advancement in production methods and transportation, but exogenous forces such as invasion and colonialization can also clearly play a major role. Both structural changes in turn contribute to a number of other structural changes, the most notable being urbanization.

One major result of these changes is an increase in the scale of the arena of interaction. Commercialization based on industry creates incentives for migration into urban areas, where large groups of individuals are put to work in factories. Commercialization based on agriculture generates more limited but similar patterns since urban areas serve as the conduit for trade between the producers of different products. Political consolidation leads to the creation of government bureaucracies as well as institutions of higher education for supplying them with personnel. This in turn creates incentives for individuals from various primordial communities to move into urban areas to compete for advancement within these institutions. Furthermore, it creates incentives for them to compete as representatives of their communities for the economic goods being allocated by the government.

Once groups individuals migrate into urban areas and large-scale economic and political institutions, competition will breed both cooperation and conflict. The reason for this is that the desired goods on offer are likely to be scarce and are apt to be distributed unequally among different groups on the basis of their effective exercise of power. On the mass level, there will be competition for factory jobs, residential

---

91 To use Anderson’s term; see Anderson 1983. The distinction also corresponds closely to that made by Patterson between “existential” and “ethnocentric” solidarity. See Patterson 1977, 43-44.

92 For a discussion of the effects of state formation in promoting large-scale ethnicity formation, see Lemarchand 1983. For discussion of the effects of urbanization, see Nnoli 1989, 21-33; Young 1976, 196-98.
space, and government services. On the elite level, there will be competition for places in educational institutions and for bureaucratic sinecures. In each of these cases, there are incentives to divide up into large factions whose members act to favor one another in the battle for economic goods.\footnote{For a discussion of this phenomenon, see Abernethy 1969, 106-7.}

The reason for this is clear: atomized individuals cannot compete very effectively against organized groups in the competition for wealth and power. Within elite bureaucracies, there are incentives for individuals to form factions whose members mutually favor one another in promotion decisions. When governments allow for political activity, there are incentives for the masses (factory workers, petty traders, and so on) to organize political lobbying groups or formal parties that pursue their mutual interests. In many cases, elite factions will be formed first, after which the leaders of these groups will attempt to form a mass following around them.\footnote{Once such a pattern of action is begun, it can be sustained as a Nash equilibrium, i.e., one where no individual has an incentive to change their behavior. Individuals will have a self-interested reason to help group members over nongroup members since the former will be more likely to use the increased resources resulting from this decision to help the individual in return.}

In the process of forming factions, the decisions faced by individuals during the coming out process can be modeled as twofold: the first is what sort of larger group to join and the second is whether to act cooperatively or freeride once one has joined such a group. In a situation in which no large-scale grouping has yet formed, there may a number of entrepreneurs suggesting various ways in which large-scale groups can be formed, and individuals can be expected to choose the groupings that provide them with the highest expected utility. This in turn will be related to the size of each potential group, the costs of organizing it, and the amount of resources that will have to be expended (for monitoring or sanctioning) to ensure that individuals within it act cooperatively.\footnote{See related discussion in Hechter 1987b, chapters II and III.}

Accordingly, it can be posited that the expected utility gained from membership within a group will be an increasing function of the size of the group, up to some ideal point that is about half the size of the entire population. This can be justified by pointing to the fact that a group that has sufficient power to dominate other groups will likewise dominate access to resources, hence it has no incentive to include more members who will simply increase the number of individuals among which such resources must be distributed.\footnote{This is a rough analogue to Riker’s theories about minimum winning coalitions. See Riker 1962.}

It can also be posited that the costs of organizing such a group will be prohibitively high unless there is a relatively simple set of criteria that can be used to define the group’s membership. This can be justified by noting that any large-scale group will have to organize individuals who basically have no acquaintance with one another. Because of
Choosing an Identity

this, there must be some impersonal criterion by which individuals can
distinguish members from nonmembers. Furthermore, this criterion
must be simple enough so that individuals can estimate the prospective
size and membership of such a group in order to rationally make their
own decision on whether to join.

Finally, it can also be posited that the resources that must be
expended to ensure cooperation in such a group will be an increasing
function of the amount of expected net selective incentives that
must be provided to individuals within the group to make them act
cooperatively, i.e., the expected sum of the private benefit the group
supplies to an individual for cooperation and the punishment meted
out by the group for failing to do so.\footnote{For a related discussion, see Becker 1983.} This is a classic collective
action problem, and it suggests that factors that raise the utility of
cooperation prior to the application of selective incentives will increase
the “efficiency” of a group as a collective enterprise.

One factor that can clearly lower the cost of cooperation is
altruism among some subset of group members. For any individual, the
net utility of cooperation vis-à-vis noncooperation, prior to selective
incentives, will be the intrinsic cost of cooperative action minus the
one’s own benefit from the increase in group welfare that occurs.
Supposing that cooperative action will benefit all members of the
group equally, this will be $u_i + a_c u_c = Du_g/n + a_c Du_g m/n - z =
(a_c m + 1)Du_g/n - z$, where $z$ is the cost of cooperation, $Du_g$ is the
increase in the aggregate group welfare that results from cooperation,
$n$ is the total size of the group, and $m$ is the number of members
of one’s own primordial community included within the group (not
including oneself). $Du_g/n$ is the net utility an individual gains from
personal benefit due to cooperative action, while $a_c Du_g m/n$ is gained
from the benefit to fellow members of an individual’s community within
the group. If the total $(a_c m + 1)Du_g/n - z$ is less than zero, then
expected net selective incentives provided by the group for cooperation
must exceed the negative of this amount in order for cooperation to
occur.

Assuming that the most efficient outcome is one in which all
group members act cooperatively, the total expected benefit for each
individual from participation in such a group if will be $(1 + a_c m)(Du_g -
r) - z$, where $r$ is the amount of resources expended per member to
ensure cooperation. $r$ in turn is a function of $\max((a_c m + 1)Du_g/n -
z, 0)$, and $r' > 0$.\footnote{Whether or not to include an individual’s altruism toward community
members with regards to $z$ in the equation will depend on whether the cost is
related to economic goods or some other goods. Within this theory, it would not
make any difference for an individual’s decisions.} What this indicates is that a higher aggregate
altruism toward members in a group (represented by $a_c m$) will increase
the expected utility of group membership in two ways: it will directly
increase the utility each member receives from the benefits received by
others in the group and it will reduce the resources required to ensure
cooperation by group members. This in turn implies that the expected utility for each individual for group membership will be an increasing function of the number of fellow primordial community members within the group. This in turn implies that, ceteris paribus, it will be more efficient to organize groups around criteria that envelop traditional communities rather than those that crosscut them.  

As noted, however, such groups will also require relatively simple criteria for membership and will have to encompass a much larger group than a single primordial community. However, the only criteria that fill such requirements and encompass entire primordial communities are primordial ones such as language, religion, customs, race or region of origin. Thus, rather than the inherent emotional drawing power of these types of factors, will be used to account for the initial formation of large-scale ethnic groups based upon primordial criteria: given existing patterns of altruism toward primordial communities, they are efficient ways of organizing individuals for collective action.

Structure and the Selection of Primordial Criteria
This notion of primordial criteria being “selected” can help explain why certain primordial criteria become the salient boundaries for ethnic groups while others are ignored. It does not fully explain, however, exactly which primordial criteria will be chosen as the basis for group membership. As noted throughout, there is no single ascriptive characteristic that inherently determines ethnic boundaries; the criterion or combination of criteria that is adopted for collective action depends on the specific structural conditions.

First of all, there will be number of potential primordial criteria that may be ruled out because they do not lead to groups of sufficient size. Second, others may be ruled out because they crosscut rather than enclose existing altruistic ties to primordial communities.

Finally, among those that remain, the one that is chosen will depend on the particular circumstances surrounding structural change. Specifically, the type of collective action that will raise a group’s welfare will depend on the shared economic and political characteristics of group members. If members of a group share some of these circumstantial characteristics, they may provide members with common objectives and interests that facilitate collective action. Although there are obviously a variety of ways in which shared characteristics can affect the logic of collective action, they can be divided into two broad categories.

---

99 Another reason may be that fellow community members can more easily monitor one another, but it is possible that community members will not be placed closely enough together to monitor each other, hence this cannot be used as a general explanation for ethnic group formation.

100 This implies that economic class at times be the basis for ethnicity formation. A class can become “ethnicized” if there is little inter-class mobility, hence causing class to encompass primordial communities. Hence, the argument would not be that class and ethnicity are opposing forms of group formation, but that class-based groups are simply one type of ethnic group.
Choosing an Identity

The first consists of placement in separate institutional channels for allocation. With regard to commercialization, this corresponds primarily with employment in different economic sectors, while with regard to political consolidation it corresponds with placement in separate political administrative units. The second category of shared characteristics consists of the uneven ability to gain benefits under current system, or rapid changes in the levels of these abilities. With regard to commercialization, this corresponds primarily with differential levels of skills for employment, while with regard to political consolidation it corresponds with de facto discriminatory policies exercised by the government.

These categories of circumstantial characteristics include many of the those discussed by conventional rational choice theorists, such as occupation (in the first category) and income (second category), which are the primary circumstantial characteristics found in the theories of internal colonization and split labor markets, as well as time of entrance into the labor force (first category) and upward mobility (second category), which are the primary characteristics found in the theory of modernization and competition. They also encompass the two main sources of mobilization noted by “politicization of ethnicity” theorists. As they claim, primordial characteristics themselves can become “circumstantialized” if they are designated by governments as a basis for distribution and/or discrimination.

In many cases, political systems will be more responsive to efforts along circumstantial lines than directly along primordial lines because redistributive policies may be carried out at lower cost by indirect means (subsidies for certain industries or regions, low prices for basic goods or high prices for luxury goods, or laws favoring strong labor unions) rather than direct means (quotas or direct transfers). When action can be directed in such a manner, they can often be expected to have higher expected marginal collective benefits resulting from cooperative behavior. This will increase the expected benefits from group membership \((1 + a_m)(Du_q - r) - z\) directly through an increase in \(Du_q\) as well as indirectly, by decreasing \(r\).

Hence, shared circumstantial characteristics, like shared primordial characteristics, will increase the efficiency with which cooperation within a group can provide benefits to group members. However, as with primordial characteristics, there will be a number of different ways in which individuals can be divided into groups according to circumstantial criteria. The exact effect of each type of characteristic will depend on the nature of the political and economic environment as well as the prior distribution of primordial characteristics among communities.

---

102 The most common case of this is the choice of official language, although this will not occur if the language chosen is not indigenous (such as English or French) or the native tongue of only a small portion of the population (such as Bahasa Indonesia).
Clearly, it is beyond the bounds of any theory to precisely model the various incentives for collective action created by every possible shared set of circumstantial characteristics. However, it can be posited that where a number of different primordial characteristics can form the boundaries for large-scale ethnic groups, the ones chosen will be those that correspond to at least one major set of shared relevant economic or political circumstances from those listed above. This, along with the preceding analysis, can be summed up in the following hypotheses:

**Hypothesis 1**: Large-scale ethnic groups will first form in societies where the majority of the population still resides in primordial communities, but where commercialization and political consolidation have proceeded to the point where many individuals will migrate from these communities to urban areas in their lifetime.

**Hypothesis 2**: In each region where such groups are formed, their boundaries will meet the following criteria:

2.1. They will be of sufficient size to comprise a substantial portion of the population within a metropole, but will not encompass significantly more than half the entire population.

2.2. Because of existing altruistic ties to primordial communities, they will encompass and not cross-cut the boundaries of these communities.

2.3. Their boundaries will align with at least one major set of circumstantial economic and political characteristics that promote cooperative action, i.e., either placement in separate institutional channels for allocation or differential ability to obtain benefits under the current system.

Of course, it is possible that more than one boundary meets the above criteria, or that none at all do so. In former case, it can be specified that the chosen boundaries will be taken from among the set which meet the criteria. In the latter case, predicting the boundaries of ethnicity will be more difficult. In either of these two cases, it can be argued that the likelihood of a coherent ethnic group forming will be smaller than in the case where a single set of boundaries stand out in meeting the criteria. When more than one boundary meets the criteria, we can expect competing political entrepreneurs attempting to organize cleavages along separate boundaries, with resulting instability in coalitions of primordial communities. When a boundary meets the criteria, competition may be unorganized or it may settle along class or occupational lines rather than ethnic ones.

Nonetheless, whatever the case, the theory does have a strong assertive component because it accounts for the timing of ethnic group formation and significantly reduces the set of potential ethnic boundaries.

It is obvious that this theory of ethnic group formation lumps together a diverse range of phenomena. Reductionism, unfortunately, is inevitable given the generality of the theory and the complexity of ethnic reality. At the same time, the theory hopefully captures a common dynamic shared by all these phenomena, in which an individual’s life process involves a “coming out” from smaller, primordial, community-based spheres of interaction to larger, urban
spheres where community-based units of cooperation are too small to effectively promote their welfare. Unlike some of the classic social theorists, I do not contend that such coming out leads to the dissolution of communal, affective ties. Instead, I attempt to show how it can lead to the expansion of such ties and the formation of large-scale ethnic groups.

Eventually, urban ethnic group leaders will go back to the rural areas, attempting to enlist their nonmigrant fellow primordial community members in their factions. The success of such efforts will be greatly promoted by the existence of a viable democratic system, which provides incentives for the rural masses to organize in ways that will be effective at the national level as well as whether the proposed ethnic groups are large enough to serve this purpose.

What this indicates, as a number of authors have noted, is that the creation of ethnic groups is the product of entrepreneurs who are attempting to “build” groups, but it also indicates that in order to be successful such entrepreneurs must shape their ideas around the existing raw material of primordial altruism and shared circumstantial interests.

The Expansion and Transformation of Group Altruism

According to the above argument, the initial basis for the creation of large-scale ethnic groups combines primordial altruism factors with circumstantial instrumental factors. However, individual identities at the point of group creation will still be directed at their primordial communities rather than the ethnic group as a whole. Ethnic altruism itself will not yet have been formed.

The logic of the argument and the coherence model, however, implies that formation of a larger ethnic altruism will eventually follow the creation of a large-scale ethnic group. A pattern of cooperation, once it begins, sets in motion a dynamic that causes internalization of altruism within the boundaries of that pattern and reinforces the resultant group’s relative efficiency in providing utility to its members. Once an individual has begun cooperating within a larger ethnic group, the same link between cooperation and altruism described for primordial communities will generate broader altruistic preference formation within ethnic groups. In particular, because cooperative interactions within ethnic groups may not always be with primordial community members, it will be doubt reducing for the individual to acquire altruism toward members of an ethnic group as a whole rather than simply toward his or her community of origin. This in turn will

104 For an early expression of this view, see Geertz 1963.
105 See Abernethy’s discussion of ethnic “missionaries” found in Abernethy 1969, 107-8. For a discussion of urban-rural links of migrants in developing countries, see Imoagene 1974.
107 For a similar statement, see Gellner, Nations and Nationalism, 49.
not clash with altruism toward community members since they are also members of the ethnic group. The acquisition of ethnic group altruism will allow an individual to continue rationalizing his or her past actions in favor of her narrow primordial community. Hence, community altruism will gradually be transformed into ethnic altruism.

This preference formation will not be hindered much by the “shadow of the past” (see chap. 3). Since members of an individual’s community of origin are a subset of the individual’s ethnic group, past cooperative activities within a community can be cognitively “redefined” as cooperative activities within the ethnic group as a whole as long as they did not involve conflict with members of the ethnic group who were not members of the community. In that case, any action that provided reliably positive amounts of economic goods to community members can be seen as an action that provided equal reliably positive amounts of economic goods (on the aggregate) to ethnic group members. Hence, if we represent an individual’s new utility function with regards to economic goods as 

\[ u_i + a_c u_c + a_g u_g, \]

where \( a_g \) is altruistic weighting toward his or her ethnic group and \( u_g \) is its aggregate welfare, increases in \( a_g \) to reduce doubt for cooperative action within the group can be obtained without increasing doubt for past actions within the community of origin simply by reducing \( a_c \) the equivalent amount.

Once this identification is transferred, moreover, it will promote further cooperation within the ethnic group by increasing the expected net utility from cooperative behavior as well as thereby decreasing the resources required to ensure cooperation, hence improving its efficiency as an organizational unit. This will in turn hamper cooperation along alternative group boundaries since that would place group members who are altruistic toward one another in conflictual situations. Hence, even if circumstantial characteristics change so that alternative boundaries provide a higher \( D u_a \) (even ethnic ones that encompass the original communities), they will be counteracted by the higher \( a_g \) that can be obtained within the existing ethnic boundaries. This indicates that after ethnic group formation, the boundaries of ethnic groups will to a certain extent be “cemented” into place by the development of ethnic altruism.

While ethnic groups will be based upon an intersection between primordial and circumstantial characteristics, there are good reasons to expect that large-scale groups will tend over time to emphasize shared primordial characteristics rather than shared circumstantial characteristics as their basis for membership. Individual primordial characteristics (as the name implies) rarely change. However, circumstantial characteristics, and hence circumstantial commonality among individuals, can be rather unstable over time. Furthermore, economic and

---

108 Although not so that \( a_c < 0 \), or else it will generate doubt for past actions within the community by creating the possibility that the individual could have had higher utility by engaging in cooperative actions with non-community members of the ethnic group.
political changes can significantly increase or reduce the circumstantiality of particular shared non-ethnic characteristics, sometimes rather quickly. Hence, groups based primarily on circumstantial characteristics will often be faced with considerable membership turnover or may discover that they no longer possess shared interests.

Among the children of urban migrants, a new generation may grow up outside local communities in a milieu in which cooperative interaction takes place primarily within the boundaries of large-scale ethnic groups, hence such ethnicity will become such an individual’s initial form of identification. This in turn will promote further cooperative interactions along ethnic boundaries. This implies that once large-scale ethnic identities have been formed major changes in their boundaries will be rare, barring some major change in patterns of migration, production or state-formation. Changes in the types and levels of shared circumstantial characteristics within these groups are hence likely to be manifested in rising and falling levels of mobilization within existing group boundaries rather than by changing group boundaries.

Nonetheless, marginal adjustments in boundaries may occur over time for a number of reasons. For instance, ethnic groups can shift to incorporate a few more communities without decreasing the group’s efficiency, depending on the past history of conflict between old and new group members and providing that the existence of a relatively simple criterion for group membership can be found that includes the new communities but does not exclude those already within the ethnic group. On the other hand, expulsions of individuals or communities will be more difficult because the past history of cooperation within the group will have created altruistic ties that will hamper future conflict.

One cause of later major changes in ethnic altruism boundaries will be changes in migration patterns. In such a case, members of existing large ethnic groups may move into new areas where the size of their group is too small for effective competition with other groups already present. In that case, there will be an incentive for them to merge with other groups in the area who face similar circumstances. The criteria for group boundaries in these “superethnic” groups will be analogous to those mentioned earlier, with the chosen boundaries not only being of sufficient size but encompassing existing ethnic boundaries and conforming with major political or economic circumstantial characteristics.

Much of the reasoning presented in this theory is consistent with the explanations of existing rational choice theories of ethnicity. Unlike such theories, however, the theory described here provides a broad microlevel explanation of why individuals will have incentives to organize along ethnic lines. Furthermore, it provides an explanation for the boundaries of ethnic groups. Rather than taking ethnic groups as

109 Not discounting, of course, the possibility of two-layered identification, one type of altruism directed at the entire ethnic group, and another one directed solely at some subset of the group based upon primordial communities.
The Origins of Ethnic Identity and Collective Action

Finally, it shows how collective action can be the cause as well as the effect of ethnic altruism.

Case Studies
The hypotheses outlined in the previous sections can be applied to actual cases in order to test their predictions about the boundaries of ethnic group formation against actual outcomes. Each of the brief case studies presented here covers a prominent instance of ethnic group formation during this century. In each case, the predictions of the hypotheses are compared to the outcomes with regards to the timing and boundaries of ethnic group formation.

The operational criterion for ethnic group formation is the initiation of significant political mobilization within a particular set of primordial boundaries where no previous history of such mobilization exists. Given the difficulties of quantifying “significant political mobilization,” comparisons must be based upon approximate indications of whether and when it occurred as well as the group of individuals it encompassed. Nonetheless, as will be shown, empirical patterns are clear enough so that it is usually possible to make noncontroversial determinations based upon the evidence present in prominent historical examinations of each case.

Clearly, even within this century, there are too many historical cases that might be classified as “ethnicity formation” to be examined within the relatively small space available here, and data on the entire set of cases (however demarcated) are not available in easily usable form. On the other hand, any manageable set of cases will be open to the criticism of possible selection bias. Nonetheless, I have tried to minimize these criticisms by trying to examine the largest and most prominent cases of 20th century ethnic group formation mentioned in the literature on race and ethnicity.

Each of these cases refers to a large population of individuals sharing specific primordial characteristics that first engaged in significant collective political mobilization during this century. Hence, it excludes the many ethnic groups with long histories of solidarity as well as those groups (e.g., the Coloureds of South Africa or the Asian-Americans of the United States) that exist as official classifications but have yet to exhibit significant collective political mobilization.

Pan-Malay Ethnicity in Malaysia
Prior to the onset of colonialism and urbanization, the residents of the Malay Peninsula could not be said to have had a high degree of ethnic identification or mobilization.

While they for the most part shared similar languages and a common religion (Islam), loyalties remained directed toward local states and their monarchs.\footnote{Nagata 1981, 97-100.}
As predicted by the theory, a politicized pan-Malay ethnicity only began stirring with British colonialization and the beginnings of migration into the Straits Settlements, the urban centers of colonial rule. Within the Settlements (Penang, Malacca, and Singapore), natives of the peninsula mixed with other Southeast Asians, Thais, Burmese, Arabs and especially with large numbers of Chinese and Indians who had been imported as a labor force by the British.\footnote{111} The existence of these large groups, particularly the economically successful Chinese, created worries among indigenous elites that they would be overwhelmed and dominated. This led to a great expansion in pan-Malay political activism.\footnote{112}

One of the preconditions for this activism was a clear definition of the boundary of Malayness, and this triggered considerable debate among Malay intellectuals.\footnote{113} The theory predicts that the boundary would be drawn in a way that would encompass existing loyalties to primordial communities and local states. Given this, a number of criteria could be used to construct such boundaries: religion (Islam), language, appearance, and family history. More concrete questions included whether to include as Malays Muslims of nonpeninsula descent, such as Javanese (Jawi peranakan) or Bugis on the one hand and Arabs and Indian Muslims on the other.

The theory also predicts that the chosen boundary would create a group large enough to form a significant but not dominant portion of the population. Because Chinese and Indians already formed a dominant portion of the population in the Straits Settlements,\footnote{114} it hence predicts that boundaries would be drawn in a very inclusive manner there. This is indeed what seems to have occurred, as the accepted definition of Malay in Penang and other Settlements came to include not only only Javanese and Indian Muslims, but also Muslims of Arab descent.\footnote{115}

However, it has not been expanded to include Chinese Muslims, despite their willingness to adopt the local language and customs. The explanation Nagata gives for this is that the Chinese “economic dominance in the country at large . . . alienates them from the rest of the Muslim community.”\footnote{116} This outcome follows the predictions of the theory for two reasons: First, the absolute number Chinese Muslims was small, hence their addition would not do anything to add to the ability of the newly-constituted Malay group to compete against other groups. Second, the inclusion of the Chinese would have blurred the alignment of the Malay ethnic boundary with the circumstantial boundary around those with relatively economic disadvantaged positions, and hence it would have hampered the Malays’ ability to act collectively to obtain common goals.

\footnotesize
\begin{itemize}
\item[111] Ongkili 1985, 5-8.
\item[112] Roff 1967, 207-10; Nagata 1979, 47-48; Young 1976, 121-25.
\item[113] Roff 1967, 242-45.
\item[114] For a large-category breakdown over the years, see Ratnam 1965, 2.
\item[115] Nagata 1981, 105-7
\item[116] Nagata 1981, 107.
\end{itemize}
On the other hand, the boundary of a pan-Malay ethnicity has not been uniformly inclusive in all parts of the peninsula. In Kelantan, where fewer Chinese and Indians are present, the definition of Malay has been much more constrained, clearly excluding those of Arab descent. This again confirms the theory, which predicts that inclusive boundaries in Kelantan would form a coalition that was much larger than that needed for political domination, hence diluting the per-individual benefits that could be gained from such domination.

In the post-independence state of Malaysia, Chinese and Indians comprise over half the population of the country. For that reason, one would expect that boundary definitions would follow along the broader lines of the Straits Settlements rather than the narrower ones of Kelantan. This has indeed been the case, and the administrative category of bumiputra, which is the basis for the affirmative action programs that have been in place since the early 1970, includes not only non-Chinese Muslims, but also non-Muslim peripheral “tribes” (orang asli). This is generally thought of as a way in which the government seeks to maximize combined pan-Malay strength. Together, this provides them with similar political strength to that of the Chinese and Indians, at about half the population.

At any rate, the strength of pan-Malay solidarity has been demonstrated by the domination of the Malay vote by the United Malays National Organization (UMNO) in the post-independence era. Even the only exception to this domination, the Parti Islam (PI), supports the previous analysis because the only states where it has been able to gain power over local government have been Kelantan and Kedah, where Malays form an overwhelming majority of the population. Patterns of electoral politics, like those of ethnic definition, follow along the lines predicted by the theory.

Pan-Igbo Ethnicity in Nigeria

Prior to colonialism, large-scale mobilization and a sense of common identity was absent among the Igbo-speaking peoples of what became Eastern Nigeria. Political organization usually did not extend beyond the village level, and what larger distinctions existed were between Igbo subgroups such as the Nri and Aro or between mainlanders (who were simply called Igbo) and the riverine Olu. As the theory predicts, large-scale ethnicity became a salient political factor only after colonialization generated political consolidation and rapid migration into large urban areas such as Lagos and Ibadan.

---

119 The formation of the UMNO splinter group, Semangat ’46, is troublesome in this respect. However, its political doctrines are practically identical to that of UMNO, suggesting that its formation reflects a failure of UMNO internal organization than any fragmentation of Pan-Malay solidarity. Furthermore, it has been unable to challenge the dominance of UMNO among Malays, and this in turn seems to be leading to a slow reconciliation process.
120 Nnoli 1989, 22; Olorunsola 1972, 7-9.
in the western region of Nigeria; Kano, Zaria, Kaduna and Jos in the north; and Port Harcourt, Onitsha, Aba and Enugu in the east. Each of these cities experienced rapid increases in population during the colonial and post-colonial era, and ethnic composition of migrant populations has not been homogeneous. Because of generally poorer agricultural conditions and inheritance customs based upon primogeniture, easterners (predominantly Igbo speakers) comprised by far the largest number of individuals who migrated out of their home regions.

However, the mere existence of migration cannot explain why ethnic group formation took place along lines of linguistic origin rather than alternative lines, most notably religion. On a national scale, both of these criteria divide the population in such a way that each group formed is significant in size but not dominant. However, the theory provides a number of explanations why linguistic boundaries were chosen over religious ones: First of all, religious boundaries tended to crosscut primordial communities. Igbo-speakers, though largely Christian, were often divided within villages between Catholics and Protestants. Likewise, Yoruba-speaking villages were similarly divided between Christians and Moslems.

Second, linguistic boundaries clearly conformed with differential levels of educational and economic advancement. From the 1930s on, Igbo speakers were more prone than others to take advantage of educational opportunities and neighboring villages were caught up in very intense competitions to “get up” economically and politically. This, along with their propensity to migrate, often brought Igbo speakers into communities where they were not only a minority but were viewed as upwardly mobile compared to non-Igbo speakers. Hence, unlike religion, linguistic boundaries aligned with a clear circumstantial boundary that created incentives for ethnic organization.

The theory predicts, however, that this ethnic boundary would initially become significant only within those areas where the group it created would form a significant but not dominant portion on the population. And in fact, pan-Igbo groups became significant only within the major cities outside their home eastern region in which Igbo-speaking migrants formed a significant but minority portion of the population. It is precisely in these locations that the group bounded by the Igbo language family would meet the characteristics for

121 For a discussion of migration patterns, see Green 1974. For a discussion of the reasons for Igbo migration, see Nzimiro 1965, 50-54.
122 It should be made clear that the eventual boundaries of the egbuc group were based upon descent from traditional Igbo language speakers rather than actual linguistic capability since many second-generation Igbo in the West and North could not speak it fluently.
123 For example, a survey in the early 1970s showed that a sample Igbo community of 1878 people consisted of 52.3 Roman Catholics and 46.5 Protestants. See Iro 1980, 185.
124 Laitin 1985, 289; see also Laitin 1986.
125 Abernethy 1969, 68-69.
appropriate size. There, pan-Igbo political mobilization was common, usually in the form of cultural/political unions,\footnote{Abernethy 1969, 110-11.} often in opposition to Yoruba mobilization in the west and Hausa-Fulani mobilization in the north. In fact, urban conflicts between pan-Igbo groups on the one hand and Yoruba or Hausa groups on the other have been primarily responsible for major outbreaks of ethnic violence in Nigeria during the colonial and post-colonial era.

Eventually, nonetheless, this type of urban pan-ethnic mobilization was reflected in the formation of political parties and the struggles between them, and gradually came to assume a national significance. For instance, in the 1959 election held just prior to independence,\footnote{In the second republic, draconian electoral engineering created strong incentives against ethnic-line voting. Even stronger engineering measures were imposed by military strongman Ibrahim Babangida prior to the aborted elections for the third republic.} the National Council of Nigeria and the Cameroons captured every seat in Igbo-majority areas, while the Action Group was nearly as dominant in Yoruba-majority areas. In the north, the Hausa-Fulani-based Northern People's Congress won over three-quarters fo the seats available.\footnote{Post 1963, 358-75.} By the 1960s, the strength of pan-Igbo mobilization had spread to the extent that the Biafran separatist war was fought in the name of Igbo self-determination and was largely supported by the Igbo population and opposed by the non-Igbo elements of the seceded eastern region.\footnote{Young 1983a, 204-11.}

Although this section has focused on the formation of pan-Igbo ethnicity, surveys of the evidence show that the theory can provide insights into group formation among the other two main ethnic groups of Nigeria. The Yoruba were the most urbanized group in West Africa prior to colonization and, unlike the Igbo, shared a sense of common origin prior to the colonial period, tracing their ancestry back to the mythical Odudua. However, they did not consider themselves a single ethnic group and in fact did not have any common term with which they used to refer to themselves. The term Yoruba originally referred only to the residents of the Kingdom of Oyo and gradually came to refer to be adopted more inclusively only after colonialization and greater contact with groups not sharing the common descent myth and language.\footnote{Nnoli 1989, 23.}

This is accordance with the predictions of the theory since such contact would create the need to band into larger groups than those based upon village or ancestral town.

Prior to colonization, Hausa-Fulani identity likewise was directed primarily at the existing large emirates rather than a common community. As the theory predicts, the formation of a pan-Hausa ethnic group was originally triggered among migrants to southern cities, where they were only able to form a viable political force when
Although religion played a stronger role as a unifying force in the north than it did elsewhere (incorporating subgroups that were not in the Hausa language family), a sense of Hausa ethnic identity is still stronger in Nigeria, where Hausa comprise a large but minority portion of the population, than in Niger, where they comprise a dominant majority.

The Luba-Kasai of Zaire

Luba-Kasai refers to a group of people originating from Kasai who emigrated to various other areas of the Belgian Congo in the early twentieth century. Luba refers to their common language family, although there are a variety of dialects. The Luba-Kasai are often referred to simply as "Luba," although this usage is somewhat confusing given the presence of other large groups, such as the Luluwa (sometimes spelled Luwa) and the Luba-Shaba, who also speak Luba dialects.

The formation of the Luba-Kasai as a politically salient ethnic group is regarded as perhaps the most extreme case of "unnatural" ethnic mobilization in Africa since it was not based on a single clear primordial criterion, but rather on a combination of linguistic and geographical criteria. Furthermore, prior to the period of migration, there was little awareness of clear differences between Luba-Kasai and Luluwa, much less any reason to make it the basis for political mobilization.

The Luba-Kasai identity began to form during the period of migration, when large numbers of Luba-speakers from the southern regions of Kasai migrated into different areas of Kasai as well as into Shaba. As migrants with little claim to land, they had a greater need to learn new skills in order to survive, and they entered mission-run educational system as well as the commercial labor force, in numbers disproportionate to their population.

As the theory predicts, however, ethnic distinctions only became politically salient with urbanization in Kasai in the 1950s and the movement of various Luba-speaking groups into metropoles, particularly Kananga (formerly Luluabourg). Since Luba speakers made up a dominant portion of the population in Kananga, the theory predicts that they would divide up into political factions based upon smaller units which would not crosscut up primordial communities. Because there were no clear-cut differences in appearance or religion between the Luba speakers, region of origin was the most obvious basis for division. However, there were numerous ways in which regionally based criteria could be devised.

---

132 Young 1976, 278.
133 Miles and Rochefort 1991.
135 The causes of migration were linked to the vulnerability of their subregion to slave trade as well as to the better opportunities available in other regions. See Young 1976, 176.
The theory predicts, though, that the regionally-based boundary chosen would align with a clear circumstantial boundary. In this case, the most obvious such boundary was that between the home regions of the economically and educationally advantaged groups that had migrated early in the century (the Luba-Kasai) and the home regions of those which had not (the Luluwa). The notion of Luba was narrowly constructed, excluding those from Luba-dialect groups who had not become modernized earlier such as Bakwa Anga or Luluwa.\footnote{Jewsiewicki 1991, 335.} Hence, its predictions accord with the boundaries that actually arose.

The ethnic groups formed by these boundaries were first manifested political with the formation of the Luluwa-Frères organization, which was devised for the purpose of defending the interests of Luluwa in light of Luba-Kasai dominance of educational and government institutions. This in turn led to a response by the Luba, and the formation of the Mouvement Solidaire Muluba (MSM).\footnote{Turner 1972, 221-22; Mukendi wa Meta 1985, 149, cited in Roosens 1989, 119.} The demands of the Luluwa-Frères gradually became more radical, culminating in a demand for the expulsion of all Luba-Kasai from Luluwa rural areas. Shortly afterward, Luluwa mobilization took an even more extreme and form, with violent attacks being directed by Luluwa against Luba-Kasai in Kananga, which in turn led to a mass exodus of Luba-Kasai back to their home regions.\footnote{Turner 1972, 223; Young 1976, 176-77.}

A similar polarization occurred, but with different groups, in the copper belt cities of Shaba (i.e., Katanga). As in Kasai, Luba-Kasai migrants were disproportionately prone to enter institutions of higher education, and hence they gradually developed a higher standard of living than other members of the local population, including other Lubaspeakers. Because of this, the theory would predict that ethnic boundaries would be drawn in such a way as to isolate the Luba-Kasai, but to leave none of the remaining groups with a dominant share of the population. In Lumumbashi (formerly Elizabethville), the largest city in Shaba, the politically salient boundaries divided the population into three main groups: the Luba-Kasai, Luba-Shaba, and “authentic Shabans,” which was comprised of Lunda and a number of small groups from southern Shaba. This was manifested first with the formation of the Confédération des Associations Tribales du Katanga, otherwise known as Conakat, led by Moïse Tshombe, which sought to combat Luba-Kasai domination of the labor force and local politics.\footnote{Young 1965, 490-96.}

As ethnic tensions polarized, this conflict exploded into a number of skirmishes. “Authentic Shaban” nationalism eventually led to the Katangan secession crisis of the early 1960s,\footnote{Young 1983a, 201-4.} during which Luba-Kasai had to be temporarily evacuated from Lumumbashi by the United Nations in order to protect them from ethnicity-based attacks.\footnote{Young 1976, 181.}
in Kasai, the increasing saliency of Luba-Kasai ethnicity was reflected in the attempt by Albert Kalonji to take advantage of the Katanga secession to declare an independent Luba-Kasai republic in Southern Kasai.\footnote{Young 1976, 188; Roosens 1989, 121.}

The Muhajir of Pakistan

Muhajir, meaning refugee in Urdu, is a generic term used to refer to the Muslims who fled from India to Pakistan during the tumult following the partition of the two countries in the aftermath of British rule or from Bangladesh following its separation from Pakistan. As indicated by the recent genesis of their name, there was no reason for future Muhajirs to feel any sense of common identity or to engage in collective political action against other groups during the era of the British Raj. Instead, they saw themselves primarily in terms of their Muslim identity, on the one hand, or their regional identities on the other. Although the vast majority were Urdu speaking, they included Gujarati speakers as well.\footnote{See Kennedy 1991, 939.} Moreover, while most came from the Hindi heartland of Uttar Pradesh, others originated from other areas and had Punjabi or Pathan names.\footnote{Ahmed 1990, 32.}

In postindependence Pakistan, however, the Muhajirs have become known as a very coherent and politically active ethnic group, despite the fact that their very existence dates back only to the Partition. This is reflected both in the massive success of the Muhajir Qaumi Movement (MQM) in sweeping Muhajir votes in the large urban areas (Karachi and Hyderabad, in the state of Sindh) where they reside in significant numbers.\footnote{Ahmed 1990, 32-33; McDonald 1988.} It is also reflected in the massive ethnic violence that has been perpetrated in those cities in support of the Muhajir cause, often in response to anti-Muhajir attacks by Pathan refugees from Afghanistan and by native Sindhis.\footnote{Kennedy 1991, 948-54. See also Phillips 1988, 18-19.}

According to the theory, the rise of large-scale ethnic groups should coincide with large-scale migration into urban metropoles. This is indeed what has occurred beginning in the postindependence period. For example, the population of Karachi, Pakistan’s largest city, has swelled from 400,000 to over 7 million in the past four decades. This has coincided with the rise of the Muhajir ethnic solidarity as well as efforts to suppress the Muhajirs on the part of ethnic groups.

However, urbanization alone does not explain why ethnic divisions in Sindh occurred largely along lines that divided refugee from nonrefugee rather than along religious lines (between Sunni and Shia). The theory provides two explanations for the current boundary. First, the requirement for sufficient size has not been met by an ethnic divide based upon religion. Although Shia make up about 15 percent of...
Pakistan’s total population, they comprise a significantly smaller percentage in the major cities of Sindh, where the Muhajir ethnicity is the strongest. The two other major groups in those cities, the Sindhi and Pathan (the latter comprised in large part of refugees from Afghanistan), are both entirely Sunni, while the Muhajir are split between a Sunni majority and a Shi‘i minority. On the other hand, the refugee category comprises a bare majority in both Karachi and Hyderabad. Furthermore, a strong Muhajir identity has been encouraged by the fact that this boundary aligns with a clear circumstantial boundary. Muhajirs on the whole have been more entrepreneurial and economically successful than other elements of the Pakistani population. In urban areas, they have replaced the largely Hindu Sindhi middle class, which fled to India after partition. Furthermore, with Jinnah and Liaqat Ali Khan as their most notable representatives, they made up a disproportionately large share of the early Muslim League elite in the immediate postindependence government. This was particularly resented by the Sindhi, who perceived Muhajir domination as the primary cause for the administrative severing of Karachi from the rest of Sindh.

This combination of factors caused the Muhajirs to stand out from the rest of Pakistani society and made them the focal point of resentments over the political and economic travails that have plagued postindependence Pakistan. This has led over time to changes in policies that have favored Sindhis and Punjabis at the expense of Muhajirs, which in turn has caused the Muhajirs to band together and respond in kind, a response exemplified by the popularity of the MQM.

Conclusion
These four cases do not constitute a definitive test of the theory since they constitute a limited sample and because the dependent and independent variables cannot be operationalized with the precision that completely forecloses alternative interpretations. Nonetheless, they show that the hypotheses generated from the theory can be used to provide reasonably assertive predictions about the location of ethnic boundaries and that these hypotheses seem to conform with historical outcomes in some of the most notable relevant cases. Because of this, it can be seen as moderate confirmation of this initial attempt at a

149 See Harrison 1986, 269.
151 Wright 1974.
152 Harrison 1986, 282.
154 Harrison 1986, 280-81.
155 Wright 1991.
deductive theory for predicting the location of ethnic group boundaries, one based upon the more general principles of a coherence model of preference and belief formation and a rational choice model of action.
Chapter 6
Structural Change, Cultural Change,
and Civic Violence

Perhaps the single area of social science in which rational choice theories have had the least impact is in explaining long-term social, political and economic change. These topics, which have generally been studied under such theoretical labels as modernization, political development, and social change,\(^1\) have been for the most part been ignored by rational choice theorists.

While it is difficult to extract a single reason why this is so, it is useful to note that almost all non-Marxist theories of long-term change implicitly or explicitly make use of cultural change as an independent causal variable or an intervening variable between structural and behavioral change.\(^2\) Conventional rational choice, on the other hand, has great difficulty in incorporating cultural diversity or change into its analysis.\(^3\) This leaves rational choice theorists without a well-established basis from which to develop their theories of long-term social and political change other than the reductionist materialist explanations of the Marxists.\(^4\)

The emphasis on cultural change is central to perhaps the most prominent hypothesis in all of social theory, that describing the transition from “traditional” to “modern” society. It has been presented in various forms by various authors, but each puts forward the

---

\(^1\) Although the terms “modernization,” “development,” and “change” are quite similar, they have somewhat different nuances. “Modernization,” along with frequent companion “evolution,” suggests a movement that occurs in one direction relative to time, while “development” suggests a movement that is normatively desirable. These nuances give them a different flavor from the almost completely neutral “change.” In practice, the terms often tend to be used interchangeably, though many theorists have been attacked for equating the latter two with movement toward the patterns of Western industrialized societies. An additional complication is the fact that “development” has to a certain extent become a generic term for Third World studies, even for static analysis, a usage which is followed to a certain extent in this book. These terminological issues have been a constant source of debate; see for instance Whitacker 1967; Huntington 1976; Riggs 1981.

\(^2\) Weiner 1966, intro., 9; Ruttan 1991, 276. Culture will be defined here in the psychological sense as collectively held preferences and beliefs, with the beliefs not based purely on logical inference from available information. See discussions in chapters 1 and 2.

\(^3\) The only ways in which culture can be incorporated into conventional rational choice is as logically held beliefs derived from access to different types of information. As argued in chapter 1, however, this is an extremely narrow and limiting view of variations in preferences and beliefs.

\(^4\) These are generally rejected by rational choice theorists for reasons having more to do with the historical origins of the two approaches rather than any methodological incompatibility. Even rational choice Marxists, like Roemer and (sometimes) Elster, do not put their emphases on Marx’s theory of social change. See also discussion in chapter 2.
two aforementioned ideal types, associating tradition with cultural characteristics such as high levels of affect, homogenous beliefs, narrow horizons and stereotyped thinking; behavioral characteristics such as low political and educational participation and geographical inertia; and structural characteristics such as poverty, subsistence agriculture, a homogeneous division of labor, and simple hereditary authority structures. Modernity is associated with individualism, diversified beliefs, heightened awareness, and openness to change; high levels of participation and mobility; and material prosperity, industrialization, urbanization, occupational specialization, formalized law, strong public administration and a stable democracy.\(^5\)

In this chapter, I examine the best-known versions of this hypothesis, put forward in the theories of “classical” nineteenth century social theorists such as Tonnies and Durkheim, which posited that societies moved inevitably forward, along a single line, from tradition to modernity. I then move onto theorists working within the past-World War II modernization approach, showing how they have significantly modified the classical version of the hypothesis. I then examine criticisms of these theories, focusing on their accusations of unilinearity, teleology and ethnocentrism. I argue that the classical theorists were vulnerable to these criticisms, while the post-World War II modernization theorists were not. However, in addressing these problems the later theorists also removed the common deductive basis that had held the classical theories together.

I then posit that the major common weakness of theories of tradition and modernity is their lack of microfoundations. No existing theory is grounded in a single, general and assertive set of assumptions about individual actors. This shortcoming, I argue, makes it difficult to trace the precise structural factors that are supposed to lead to changes in culture as well as the precise effects of cultural change on behavior. This in turn can account for many of the attacks directed against these theories and the failure of later theories to cohere into a consistent whole.

I then provide my own theory of cultural change, one that is derived from the coherence rational choice model. It focuses in particular on the effects of structural changes such as industrialization, commercialization, and bureaucratization on human interaction patterns, and the effects of these pattern changes on individual and collective beliefs and preferences. I argue that such changes tend to lead to more diffuse and cross-cutting patterns of cooperative interaction. This in turn, through the coherence reduction process, affect the preference and belief patterns of individuals participating in such interactions. I focus in particular on patterns of altruism, arguing that socioeconomic changes of the sort listed lead to weaker and more diffuse patterns of altruism.

I apply this theory of cultural change to analyzing one particular behavioral pattern, variations over time in civil violence. This is one

---

of the more controversial topics in the overall analysis of long-term change, particularly given the initial upsurge in violence, most notably ethnic violence, that appears to accompany economic growth. I show why the relative benefits of violent action can be shown to initially rise, then fall, with increased diffusion in the altruism patterns of individuals. This in turn affects the tendency of rational actors to engage in collective violence. Hence, I show how an important long-term behavioral phenomenon can be explained by means of a simple and general set of assumptions about individual actors. This in turn can provide insights into how institutional design can affect patterns of behavior.

Theories of Tradition and Modernity
Theorizing about tradition and modernity can be split into two major stages: the first comprised of the work of nineteenth century European social theorists, while the second comprised of the work of post-World War II modernization theorists. As noted earlier, I argue that the postwar theorists introduced major changes in the nature of modernization theories, eliminating some of their reductionist assumptions but not replacing them with any alternative assumptions of comparable generality and assertiveness.

Classical Social Theories
19th century theories of social and political change were inspired by the historical changes that had occurred in the West during the industrial revolution as well as by the contrast between Western and colonized non-Western societies. Theoretically, they were inspired in part by the Darwinian revolution and its emphasis on the gradual modification of the species over time.

Two of the early analysts of tradition and modernity were Maine and Spencer. Maine’s *Ancient Law* associated traditional society with law based on patriarchal hierarchy and family-based collective responsibility, which he called *status*. Modern society, on the other hand, was associated with law based upon freely agreed upon transactions, which he called *contract*. Spencer was concerned more with the complexity of society, positing that population growth would promote a change from a simple, homogeneous “military” society to a heterogeneous “industrial” society comprised of interdependent parts. Unlike most of their successors, Maine and Spencer concentrated on describing changes in patterns of behavior, and did not attempt to explain the individual or collective psychological basis of these changes.

An emphasis on psychological and cultural factors can be traced back to Tonnies, who distinguished between the “natural will” that determines behavior in *Gemeinschaft* (“community”) and the “rational will” that determines behavior in *Gesellschaft* (“society”). Natural will reflects affective feelings between kin or neighbors, which is the

---

7 Spencer [1892] 1969
8 Tonnies [1887] 1957.
result of shared characteristics and experiences of those within a small rural community. Rational will reflects an isolated calculation of self-interest, which arises in industrialized urban areas.\(^9\) Hence, Tonnies posited a direct causal link between industrialization and cultural change. Furthermore, he viewed cultural change as reflecting a decline of altruistic preferences, the partial identification of one’s own interests with those of a particular community.

Durkheim, on the other hand, focused in particular on the sources of solidarity, the forces that hold a collection of individuals together.\(^10\) He asserted that in traditional rural villages a simple, undifferentiated division of labor creates mechanical solidarity, in which an individual’s will is subsumed into the collective conscience of a homogenous group. Population growth, however, will lead to an increased density of interactions and a more specialized division of labor. Such a society will be held together by organic solidarity, which is based on the mutual dependence of individuals, rather than any shared sentiments.\(^11\) This change will be reflected in a movement from repressive law, which exacts societal retribution on those who offend the collective conscience, to restitutive law, which is a compensatory transaction between the offender and those directly injured.\(^12\)

There is clearly compatibility in the terminology used by Durkheim and Tonnies to describe the cultures of traditional societies, with natural will corresponding very closely to mechanical solidarity. On the other hand, while their depictions of modern society are not logically incompatible, they differ in their emphases. Tonnies posits that modernity leads to the pursuit of individual self-interest, while Durkheim posits that an alternative source of solidarity can be found in the dependence of individuals on one another in a division of labor. At any rate, Durkheim and Tonnies came to reflect mainstream thought on the cultural consequences of modernity, which is connected with socioeconomic forces such as increased industrialization, commercialization and urbanization. This view of the contrast between tradition and modernity was reinforced by later well-known contributions to social theory, such as Weber’s distinction between traditional, charismatic, and rational-legal authority and his hypothesis about the progressive rationalization and bureaucratization of society;\(^13\) as well as Simmel’s analysis of the relationship between group size, patterns of interaction, and social structure.\(^14\) In each case, change was thought to be a gradual process, which in which forward movement took place simultaneously along a variety of fronts: economic, social, and political; cultural, behavioral, and structural.

---

\(^12\) Durkheim (1893) 1984, book 1, chap. 5.
\(^13\) Weber 1978, part 1, chap. 3 and part 2, chap. 11.
\(^14\) Simmel 1908, part II, chap. 2, 6.
The Postwar Modernization Approach
Even before the end of World War II, American social scientists were active in building theories of tradition and modernity along the lines of classical social theory, the most notable examples being Cooley’s investigation of primary groups\textsuperscript{15} and Redfield’s analysis of the folk-urban continuum.\textsuperscript{16} However, the theme of tradition and modernity became one the most important focuses of American social science with the postwar advent of the “modernization” approach.\textsuperscript{17}

The historical and theoretical factors behind this resurgence of interest were far different from those that provoked the classical theorists. The resurgence was inspired in large part by decolonization and the large number of newly independent states that had come into being as well as a largely cold-war-inspired concern about their political futures. It was also inspired by the memories of the Great Depression and the breakdown of democracy in societies such as those of Germany and Japan which had led to political extremism and war. The confluence of these interests created great concern over how to promote stable democracy in the newly independent states. Theoretically, the direction that this new research took was inspired in large part by the “behavioral revolution” and its associated methodologies of survey research, statistical analysis, and systematic cross-national comparison as well as the dominating influence of Parsons on the American social sciences.\textsuperscript{18}

The focus on democracy and its prerequisites created a new dimension in the analysis of tradition and modernity. Previous analysis of the structural consequences of modernization had focused on socioeconomic phenomena such as industrialization and urbanization, which could plausible by posited to move relentlessly forward with time. However, it was manifestly obvious that movement toward democracy did not occur on a smooth linear path. Not only were many different forms of democracy possible but the strength of democratic institutions could clearly stagnate or even decay within a particular country over time.

This meant that greater attention had to be paid to how the process of long-term change failed to follow a smooth line and how disturbances could take place within this process. For this type of analysis, the contribution of Parsons was important, albeit in an

\textsuperscript{15} Cooley 1909. See also Cooley 1902 and Cooley 1918.

\textsuperscript{16} Redfield 1930 and Redfield 1941.

\textsuperscript{17} As with the above discussion of “classic social theories,” this section is not meant to be a thorough review of all postwar modernization theories. Instead, it is meant to provide some examples of each major subtype of the modernization approach in order to analyze their differences and similarities. Those wishing for more complete recent surveys of this literature can look at: Sztompka 1993, chapters 8 and 9; So 1990, part I; Harrison 1988, chapters 1 and 2; Janos 1986, chap. 2. For recent compilations of articles by “big names,” see Haferkamp and Smelser 1991, which features sociologists, and Weiner and Huntington 1987. (Boston: Little, Brown and Co., 1987), which features political scientists.

indirect manner. The “Pattern Variables,” which he had introduced in his middle-period work,\textsuperscript{19} were the first prominent attempt to modify the tradition-modernity distinction so that long-term change could occur nonuniformly along multiple dimensions. Instead of a single dimension, he listed five, each of which was deemed essential to fully characterizing a pattern of behavior: Affectivity refers to the immediate gratification of desire, while affective neutrality refers to the delay of gratification. Collectivity-orientation refers to a concern for how one’s activity will impact upon others, while self-orientation refers to the absence of such a concern. Particularism refers to action based upon relationship-specific rules, while universalism refers to the action based upon generally applicable rules. Ascription refers to the tendency to judge people on the basis of inborn characteristics, while achievement refers to the tendency to judge people on the basis performance. Diffuseness refers to broad, all-encompassing relationships, while specificity refers to specialized, compartmentalized relationships.

Despite the possibility of multidimensionality, most of Parsons’ successors used the pattern variables as mutually reinforcing indicators of a single cultural pattern. For instance, Levy posited that preindustrial societies were associated with particularism and functional diffusion, while industrial societies were associated with universalism and functional specificity.\textsuperscript{20} Hoselitz associated collective orientation, universalism, achievement and functional specificity with wealthy societies.\textsuperscript{21} Sutton distinguished between agricultural society, which is characterized by ascription, particularism and diffusion, and industrial society, which is characterized by universalism, specificity, and achievement.\textsuperscript{22}

However, because they provided more than one dimension, the pattern variables allowed theorists to track uneven processes of change. Lipset, for instance explored the consequences for political behavior of different combinations of values in different variables. In his \textit{The First New Nation}, he compared the U.S., Britain, France and Germany on the dimensions of achievement-ascription, universalism-particularism and specificity-diffuseness dimensions as well as his own egalitarianism-elitist dimension. He posited that the U.S. and Germany both had values along the different dimensions that are mutually supportive, but that U.S. values were conducive to democracy while German values were not. On the other hand, Great Britain was seen to have one set of values governing economic and political interactions and another governing social interactions. Nonetheless, the former set of values were posited to be sufficient to sustain democracy. Finally, the French population was viewed as divided into different social groupings with incompatible sets of values, hence not conducive
to democracy.\textsuperscript{23} Parsons himself, when applying the ascriptive-achievement and particularistic-universalistic variables, built a matrix with four possible outcomes, each corresponding with a different set of actual countries. The universalistic-achievement, universalistic-ascriptive, particularistic-achievement, and particularistic-ascriptive combinations were associated with the U.S, Germany, China, and Latin America, respectively.\textsuperscript{24}

Even though most later modernization theories did not explicitly make use of the pattern variables, they set a precedent for a type of theorizing that did not assume that modernization was all of a single piece. This type of theorizing often focused on the possible discoordination between socioeconomic modernization and political modernization or development, though this was not the only type of comparison made. While “classical”\textsuperscript{3} theories could still be found,\textsuperscript{25} most of the modernization theories of the 1960s were concerned primarily with modifying this view.

These theories can be divided into three main categories. Each represented a different sort of break with the unilinear and progress-oriented assumptions of classical social theory. The first focused on the transitional stage between tradition and modernity, particularly under conditions of rapid socioeconomic modernization. These theories posited that certain patterns of culture and behavior manifested during such a period were unique and not simply at a midpoint between the patterns found in traditional and modern societies. The second analyzed how modernization could proceed unevenly along different dimensions and how societies undergoing modernization displayed varying patterns of culture and behavior. The third allowed for the actual breakdown or decay of modernization along certain dimensions, and the possibility that modernization along one dimension might lead to demodernization along others.

The great salutary effect of these breaks was that they allowed modernization theorists to more accurately analyze the myriad patterns of social and political change that could be found among the countries of the Third World. Rather than imposing a single pattern on all countries, it allowed for more nuanced theories that fit the individual characteristics of each case. However, I will argue that this “opening up” of the modernization approach had a cost: by removing the reductionist assumptions of unilinearity and irreversibility, it stripped the modernization approach of its only common deductive elements. Because of this, the new theories, while they had a common object of analysis, did not share any clearly stated common premises. Hence, it

\textsuperscript{23} Lipset 1963, chap. 6.

\textsuperscript{24} Parsons 1951, 101-20.

\textsuperscript{25} Most notably Lipset’s Lipset 1959b, which was later published as chapter 2 of Political Man. In it Lipset posited that socioeconomic development promotes greater democracy by causing the lower classes to develop longer time horizons and the upper classes to become less protective of their wealth. Even Lipset, however, adds the caveat that too rapid development can lead to lower class extremism. See Political Man, 68-72.
became more and more difficult to characterize them as a whole and to place them within a single model of long-term social and political change.

_Transitional Society and Politics_

The notion of transitional society politics began to make its appearance in the 1950s, becoming prominent in Lerner’s _The Passing of Traditional Society_. In it, Lerner put forward the concept of the “transitional man,” who had been made aware of possibility of a better life by the availability of the mass media and efficient transportation, yet was often frustrated because of the gap between his desires and the ability of existing institutions to fulfill them. Because of this, such a man is particularly susceptible to political extremism. Nonetheless, Lerner maintained the hypothesis that all transitional societies would in the long run evolve into societies that were modern along all dimensions.

Pye’s analysis focused in particular on the cultural characteristics of political elites in transitional societies. His main thesis was that successful political development depended on the success of elites in forming a sense of shared identity that can replace the traditional one. In addition, he described a number of behavioral and structural characteristics of transitional politics. These include the tendency of political parties to represent ways of life rather than functionally specified interests, sharp generational conflict, a lack of consensus over legitimate political institutions, and the prevalence of charismatic leaders.

Apter’s analysis of transitional authority systems provided a two-dimensional classification of authority types, revolving around the dimensions of sacred-consumatory versus secular-instrumental values and hierarchical versus pyramidal decision-making. He concentrated in particular on comparing the sacred/hierarchical type of system, which he called a “mobilization” system, the secular/pyramidal type which he called a “reconciliation” system, and the secular/hierarchical type, which he divided into “modernizing autocracy,” “military oligarchy” and “neomercantilistic” subtypes. He asserted that neomercantilistic systems, which depend on a combination of strong centralized leadership and the public justification of state policy, will be optimal for the early stages of socioeconomic modernization. Mobilization systems, which depend on raising support for the state’s policies to the status of national “religion,” will be the optimal for the later

---

26 _The Passing of Traditional Society_, 69-75.
27 Parsons 1951, 402.
28 Parsons 1951, 46-47.
30 Pye 1962, chap. 2.
31 Apter 1965.
32 Apter 1965, 22-38.
stages; while reconciliation systems will be successful once such modernity is achieved. Hence, while Apter’s typology allows for a more complex categorization of regimes that the simple traditional-modern dichotomy, his theory retains the implication that modernizing states and societies will follow similar paths, though these paths will involve nonlinear political changes.

Perhaps the most concentrated analysis of one aspect of transitional society was made by Riggs. After first making a comparison of traditional and modern public administration based on the pattern variables, he described how administration in transitional societies represented yet a third type of pattern. Using the analogies of “fused,” “prismatic” and “diffused” societies, he argued that the middle stage had its “own peculiar logic.” Its characteristics include a “bazaar-canteen” economic arena in which price inassertiveness leads to constant haggling, an “agglomerated” status system in which there are several different criteria for eminence but one overarching elite, a “polynormative” atmosphere of conflict over basic cultural norms, and a “bifocal” authority system that is formally narrow in scope but in fact has few limits.

Theorists of transitional society not only posited that the journey from tradition to modernity would not be smooth and linear, but that it could slow down quite dramatically or even be blocked if the necessary political conditions were not present. Pye and Riggs in particular were quite explicit in their doubts about the inevitability of the movement from transitional society to modernity. Pye noted that both traditional and modern elements can exist side by side and that the success of development “depends upon whether the traditional patterns tend to reinforce . . . the modern superstructure of relationships or whether they tend undermine and disrupt it.” Riggs asserted that “there is no reason why transitional societies should succeed in their efforts to become modern, nor is it inevitable that the remaining traditional systems should become transitional or modern.” Nonetheless, the theorists of transitional society concentrated on analyzing the conditions for movement forward, not movement backward. Moreover, they focused their energies on analyzing the common characteristics of societies in this stage of modernization rather than attempting to show how the stage could take various forms.

34 Riggs 1957.
35 Riggs 1964.
36 Riggs 1964, 463.
37 Riggs 1964, chapters 3-6.
38 Pye 1962, 38.
39 Riggs 1964, 36.
40 Riggs did talk about the possibility of “negative development,” but did not focus his theory on describing the conditions under which it occurred. See Riggs 1964, 117-18.
Multilinear Modernization

Moving another step away from the classical view of modernization, a number of theorists provided explicit propositions about the different paths that societies could take between tradition and possible modernity, including the conditions under which each path would be taken. Shils was the first to provide a clear typology of the different types of political regimes that could be found in transitional societies, distinguishing between political democracy, tutelary democracy, modernizing oligarchies, totalitarian oligarchy and traditional oligarchy. For each type, a set of preconditions was provided relating to the stability, coherence and effectiveness of the ruling elite; the practice and acceptance of opposition; the machinery of authority; the institutions of public opinion; and the civil order. Despite this variety, however, Shils asserted that the latter four types of regimes are viewed by leaders as second best expedients, to be replaced with political democracy if or when successful modernity is achieved.

A more complex typology of paths was provided by Black, who posited four stages of development and seven different paths that could be taken. The four stages are labeled the challenge of modernity, the consolidation of modernizing leadership, economic and social transformation, and the integration of society. The seven patterns were not given explicit labels, but are distinguished according to five main factors: the time when the transfer of power from traditional to modern leaders occurred, whether the challenge of modernity to traditional leaders was internal or external, whether the society had enjoyed continuity of population and territory, whether the society had experienced prolonged colonial rule, and whether the society already had developed institutions that could be adapted to modernity.

Another type of classification was produced by several members of the Committee on Comparative Politics of the Social Science Research Council, and focuses on five “crises” that developing countries must pass before passing onto modernity. The crises include those of identity (creating a sense of nationhood), legitimacy (creating common beliefs about the proper nature of government), participation (the incorporation of new interest groups in politics), penetration (building institutions for effective policy implementations), and distribution (building consensus about how goods are to be allocated). According to the authors of this framework, the exact development pattern shown by each country would be determined by the sequence in which each

41 Shils 1962.
42 Shils 1962, part 3.
43 pp. 48-9.
44 Black 1966.
45 Black 1966, 67-89.
46 Black 1966, 96. This classification is used to classify the development patterns of 175 industrialized and developing countries. See pp. 90-4.
47 This was introduced in Pye 1965, 62-67 and became the basis for a jointly written volume; see Binder 1971, 52-67.
of these crises was met and whether or not it was overcome, although it was acknowledged that general hypotheses would not be easy to extract. Like theorists of transitional society, multilinear modernization theorists made no assurances that change would eventually push forward into modernity. In particular, this would depend on how successfully states met the “challenges” or “crises” of development. However, they did not attempt to systematically analyze in detail the conditions and consequences of failure.

**Breakdown and Decay**

A final abandonment of classical premises can be found in theories of breakdown and decay, which explicitly analyze how certain dimensions of modernity can move in reverse over time and the conditions under which this will occur. This analysis of “breakdown” and “decay” arose with two theories presented at about the same time by Eisenstadt and Huntington.

Eisenstadt’s article focused on how continuous socioeconomic modernization could increase the increased political participation of various social groups, which in turn led to prolonged conflict between such groups in the absence of integrating formal institutions and shared values that were capable of mediating their conflicting demands. This conflict could in turn lead to the breakdown of centralized modern political institutions formed in the earlier stages of modernization and a return to more traditional ones within smaller arenas of interaction. This lack of integrating mechanisms was in turn attributed to the cultural “self-centeredness” of groups, which is a reflection of the lack of intergroup cohesion among elites, the weakness of their existing links with the masses, and their lack of internal value transformation.

Huntington’s article put forward a similar argument, suggesting that political development was dependent upon the strength of institutions to cope with the pressures created by socioeconomic change. This strength depended on the scope of political organizations and their level of institutionalization, which in turn depended on four factors: adaptability (the ability to survive in changing environments), complexity (the multiplicity of subunits), autonomy (the extent to which functioning is independent from social forces) and coherence (internal unity). In the absence of this type of strength, the tide of rising participation and expectations, could lead to instability and

---

49 Eisenstadt 1973a. This is presented in somewhat elaborated form in Eisenstadt 1966, chapters 6 and 7.
52 Huntington 1986. Much of this essay was incorporated into Huntington 1968, chap. 1.
53 Huntington 1986, 103-14.
in turn to “institutional decay,” i.e., the replacement of democratic institutions with military dictators or charismatic leaders.\textsuperscript{54}

Both Eisenstadt and Huntington made it clear that not only could certain dimensions of modernity move in reverse, but movement forward in one dimension could lead to movement backward in another.\textsuperscript{55} Hence, to the extent that modernity was considered desirable, as Packenham noted, it was no longer possible to assume that “all good things go together,”\textsuperscript{56} but rather that tradeoffs must be made between different aspects of modernity. This did not negate the fact that there appeared to be many a large group of societies (the Western industrialized ones) that displayed all the characteristics of modernity, but made their existence a paradox rather than an inevitability.\textsuperscript{57}

The cumulative effect of the later theories in the modernization approach was to unravel the classical model. Transitional society theories set the stage for the questioning of the unilinear, irreversible view,\textsuperscript{58} while multilinear theories and theories of breakdown and decay explicitly contradicted this view.\textsuperscript{59} Moreover, while these theories were a natural result of trends in the modernization literature, they brought the term modernization itself into question since the variables they were describing had values that moved both forward and backward with the passage of time. The theories dealt not with a single, forward-moving process, but with a more complex analysis of long-term change.\textsuperscript{60}

\textbf{Criticisms of Theories of Tradition and Modernity}

During the late 1960s through the 1980s, both the classical distinction between tradition and modernity\textsuperscript{61} and the postwar modernization\textsuperscript{62} approach\textsuperscript{63} came under a number of attacks from critics. These criticisms took a wide variety of forms, but for the most part they centered around the closely related themes of unilinearity, teleology, and ethnocentrism. By unilinearity, it was alleged that theories of tradition and modernity assumed that developing countries would all follow a single path of long-term change. By teleology, it was alleged

\textsuperscript{54} Huntington 1986, 116-17.
\textsuperscript{56} Packenham 1973, 288.
\textsuperscript{57} Huntington 1987, 17.
\textsuperscript{58} Eisenstadt 1973c, 17.
\textsuperscript{59} Eisenstadt 1987.
\textsuperscript{60} Huntington 1976, 60.
\textsuperscript{61} Gellner 1969, chap. 1; Bendix 1967; Gusfield 1967; Rudolph and Rudolph 1967, chap. 1; Kothari 1968; Nisbet 1969, especially chapter 6; Lauer 1971.
\textsuperscript{62} Despite reservations about whether some of the theories covered above really describe a process that can be called “modernization.” I will still use the term to refer to the approach as a whole since it reflects established usage.
that all events were interpreted as forward movement along this path, and reversability was denied. By ethnocentrism, it was alleged that this assumed path reflected the historical experiences of Western countries and was inappropriate for non-Western countries.

Embedded in these criticisms were allegations that the theories assumed that economic, social and political variables all reinforced each other in a single direction and that socioeconomic change would be free from political conflict. Furthermore, it ignored the diversity that existed among non-Western countries and the special effects of external factors (particularly Western imperialism) on change in non-Western countries.64

These criticisms can be evaluated according to two criteria. The first is the extent to which, if true, they undermine the usefulness of theories of tradition and modernity; the second is whether they are accurate portrayals of the theories. With regards to the first criterion, it can be noted that there is nothing theoretically wrong per se with asserting that all societies will follow certain patterns over time; such assertions are the essence of all general theories of long-term change. Nor is it necessarily ethnocentric to assert that Western industrialized countries have followed these patterns earlier than Third World countries as long as the earlier movers are not glorified in the process. However, the real crux of the issue is whether such theories provide useful explanations of change, and the history of developing countries in the postwar era has created major anomalies for any theory that adopts the classical view of modernization, where cultural, behavioral, and structural variables in political, social, and economic spheres are all assumed to move forward together toward a modernity defined in terms of the industrialized West. It has been shown, at the very least, that the variables associated with political modernity do not precisely track the variables associated with economic and social modernity, and that they can even move in opposite directions. Hence, judged by the first criterion, the criticisms are quite damaging.

On the other hand, the criticisms become less relevant in light of the second criteria, because they cannot accurately be applied to theories of tradition and modernity as whole, and are particularly inappropriate for the postwar modernization approach. As was shown, the entire trend of the approach in the last few decades has been away from the unilinear and irreversible assumptions of classical social theory and toward more nuanced explanations. The explanatory shortcomings of classical theories had been noted fairly early by modernizationists, as had the dangers of unilinearity, teleology, and ethnocentrism.65 Furthermore, later theories in the approach, which incorporated multilinear modernization as well as breakdown and


65 A warning by Almond in 1968 uses these exact terms. See Almond 1970, chap. 10, 287. See also authors cited in Almond 1987, 447-50.
Choosing an Identity

decay, clearly could not accurately be described as having a unilinear or teleological quality.

Why was it, then, that so many critics chose to apply these criticisms in a blanket manner to the entire modernization approach? Two plausible reasons may be put forward. The first is that the idea that all societies would follow a single, irreversible path from tradition to modernity was the defining assumption that held the classical theories together. Even after it was abandoned by many later theorists, it remained the sole proposition that could be associated with the modernization approach as a whole. Other than this, there is nothing deductive to distinguish modernizationists from any other set of theorists studying long-term social and political change.\(^66\) Hence, it is understandable, if not exactly justifiable, that critics would use this as their target when they wished to launch a single attack against the entire approach.

Secondly, long after unilinearity and irreversibility had been rejected, theorists in the modernization approach retained the concepts of tradition and modernity employed by the classical social theorists, concepts that were originally derived from the observation of long-term change in the West. Although these concepts no longer have any shared deductive significance, they still implicitly represent the polar ends of the process of long-term change.\(^67\) Furthermore, because the newer theories of the modernization approach have not coalesced around some alternative set of assumptions, they cannot put forward any clear alternative vision about how to organize thinking about change. Even for multilinear theories or theories of breakdown and decay, the classical unilinear, irreversible path remain the baseline from which diversions are measured.

**Deductive Incoherency**

Both of these reasons put a focus on the main weakness of the revised modernization approach: its lack of alternative assertive assumptions. Not only does this weakness prevent the revised approach from clearly distinguishing itself from the assumptions of the older classical approach, but it also creates a number of methodological problems not faced by the older approach.

Because the revised approach allows for multilinear and reversible paths of change, it can literally accommodate any conceivable combination of economic, social, and political phenomena as well as any type of change in these characteristics, within its range of allowable outcomes. The positive side of this is that theories within this approach can examine in a very nuanced and accurate way the complexity and distinctiveness of changes in different societies. Nothing is ruled out a priori. On the other hand, this means that, although individual theories within the approach may be falsifiable, the approach as a whole is not.\(^68\)

\(^{66}\) See discussion in Roxborough 1988, 775-77.

\(^{67}\) Smith 1973, 87.

\(^{68}\) For a version of this criticism applied to modernization theories of ethnicity, see Horowitz 1985, 101. For a defense of such an approach, see Pye and Pye 1985, 10-11.
Because they are not restrained by an overarching set of assertive assumptions, there is a tendency for even individual theories in the approach to have an empirically driven and *post hoc* flavor. Without restrictions, individual theorists are free to constantly alter the assumptions of their theories to accommodate newly observed facts. Although this might allow the hypotheses generated to closely conform to observed reality, it will also create suspicions that such conformity is spurious rather than a reflection of the validity of the assumptions and would create doubts about the reliability of their predictions about the future.69 Assertive assumptions, on the other hand, will reduce the theorist’s ability to manipulate theory to “fit the facts” and will improve the credibility of any resulting predictions.

Another result of the approach’s lack of assertive assumptions is that there is no basic consensus on the causal relationship between different variables and, barring the unlikely prospect of clear-cut empirical evidence, there is no agreed-upon way for resolving this lack of consensus. Although a number of theorists have viewed economic structural change (industrialization in particular) as the “prime mover” in long-term change70 other theorists have viewed cultural change as the precondition for economic change.71 And, as was noted earlier, there has been an even wider degree of disagreement on the relationship between socioeconomic and political structural variables. The upshot of this disagreement is that it is very hard to extract from the literature as a whole any clear set of propositions about what influences what.72 This in turn has made theories difficult to cumulate and has created a sense of theoretical incoherence. Although classical theories were not in much more agreement about the causal relationship between variables, the assumption that all variables would move together toward modernity made this a much smaller problem for theoretical coherence than it is for theories in the revised approach.

*Rebuilding the Deductive Basis*

One way to address this problem would be to retain unilinearity and irreversibility assumptions, but explicitly limit them to a specified subset of variables. Most theorists, after all, continue to hold a unilinear and irreversible view with regard to most areas of socioeconomic change and certain areas of political change. Hence, these theorists might try to make a clear distinction between those aspects of long-term change that are thought to move forward together in an irreversible manner,

---

69 See the end of the next chapter.
71 Notably McClelland 1961 and Hagen 1962. For a criticism of these types of theories, see Hirschman 1965.
and those that are not. However, there is unlikely to be consensus on the variables that should be put within a bare-bones unilinear, irreversible model, and which should be excluded. Although a number of theorists have already made a distinction between modernization, socioeconomic development, or social mobilization on the one hand and political modernization or political development on the other, few would go so far as to claim that the dividing line between unilinearity and nonunilinearity is simply that between society/economy on the one hand and politics on the other. Some political phenomena, such as expansion in participation, are still widely considered to be part of a unilinear process, while some socioeconomic phenomena, such as economic growth, can clearly be reversed. Furthermore, a reduced unilinear, irreversible model, however, conceptualized, would have limited explanatory power since it would have nothing to say about the variables that fall outside its purview.

A radically different way of rebuilding a deductive basis for theories of long-term social and political change would be to reconstruct it from the microlevel, from a general, assertive set of assumptions about individual actors. This has a number of advantages. First, it would avoid the problems of reductionism on the one hand and incoherence on the other that are inherent in any attempt to build a deductive scheme based on two opposed ideal types. Second, it would allow theorists to directly address the questions of causation between structural variables by examining the ways in which individuals react to their surrounding environments, and how this creates changes in their environments. Finally, it would avoid the need to posit that the process of long-term change stops at some point, when the “modern” end of the continuum is reached.

One way of building a microbasis, of course, would be to adopt conventional rational choice assumptions. However, this would force the resulting theories to ignore the important role that culture has played in existing modernization theories. This in turn would significantly limit the explanatory power of the resulting approach, a fact that is indirectly borne out by the present shortage of rational choice theories that attempt to cover the same areas of analysis as the modernization approach. Hence, what is needed is an approach that can incorporate culture in its analysis, but retains both generality and assertiveness. This in turn requires a model that can that can

---

73 Presumably, this former category would contain more that a single variable, otherwise the revised concept of modernization would be subject to the criticism that it is synonymous with another concept (that represented by the variable), hence superfluous. See Huntington 1976, 42-44; Tipps 1976, 68; Smith 1973, 71-79.

account for changes in both the physical environment (structure) and the cultural “environment” (patterns of preferences and beliefs).  

An Coherence Rational Choice Theory of Structural Change, Altruism, and Collective Action

The pertinent question, then, is whether a theory based upon a general and assertive actor model can incorporate and explain long-term cultural change as well as its effects on behavioral change. In this section, I attempt to show how the coherence rational choice model can in fact be used to build such a theory. The main structural assumption of this theory is that various types of technology-driven change will lead to systematic changes in patterns of economic interaction due to the incentives created by the imperatives of production efficiency. These changes in patterns of interaction, through the coherence reduction process, lead to changes in individual patterns of altruism. Finally, these changes in patterns of altruism lead to changes in political behavior, specifically patterns of political order and violence.

As this indicates, I do not attempt model all the possible effects of “modernizing” structural change on culture and behavior; these are undoubtedly much too complex and numerous to include in any single theory. Instead, I try to focus on one particular systematic effect of such changes on culture (i.e., patterns of altruism), and how it is manifested in individual and collective behavior.

Modernizing Structural Changes and Patterns of Interaction

Two particular types of changes in patterns of interaction are discussed: The first, broadening, is an increase in the total number of other individuals with whom an individual has interactions within a fixed period of time. The second, fragmentation, is a decrease in the average frequency of interactions with any particular individual. Broadening reflects the larger number of individuals that are brought into contact to engage in cooperative or conflictual activities. Fragmentation reflects the increasing compartmentalization of interactions into various specialized niches. These in turn reflect individual-level effects of the two major structural changes that have long been associated with modernization: expansion and differentiation.  

Both of these changes can be attributed to a number of the structural processes discussed in the modernization literature, which can

---

75 This implies that formalized notions of how social change operates are incomplete if they simply view change as a dialectic between structure and action, unless preferences and beliefs are incorporated as part of structure. For examples of the structure-action point of view, see Holt and Turner 1975, 991; Boudon 1986, 31-32.

76 See Sztompka 1993, 71-74, 100. One of the other changes mentioned by Sztompka, individualization, is seen here as an effect on preferences and beliefs caused by the structural changes.
be grouped under broad categories such as industrialization, commercialization, and bureaucratization. Each promotes broadening. Industrialization creates the incentives for economies of scale, which in turn create incentives for incorporating larger populations within common productive activities. Commercialization creates incentives for geographical dispersion and increased mobility in economic transactions. Bureaucratization creates encompassing institutions for governance and training that bring together individuals from diverse backgrounds. All three processes promote urbanization, which itself promotes further broadening: industrialization through the development of factories and supplier networks; commercialization through the development of market and warehousing facilities; and bureaucratization through the creation of government, military, and educational centers.

Likewise, each process promotes fragmentation. Industrialization promotes wage labor, which in turn creates differentiation between interactions at the production site and interactions within the family. Commercialization promotes specialization in production, which in turn creates numerous sites for trade in different goods, fragmenting boundaries of interaction for individuals who need to procure a wide variety of goods. Bureaucratization likewise promotes specialization in government services, with an analogous result.

On the other hand, it is hard to specify in general the extent to which each structural process will tend to promote broadening relative to the amount to which it will promote fragmentation since each process itself represents a varied set of phenomena. Furthermore, although each of these structural processes has a general tendency to move forward with time, this does not occur with sufficient certainty to justify using such a teleological word as modernization to describe any of them. Deindustrialization, decommercialization and debureaucratization all have been known to occur, not only during periods of political turmoil (e.g., during the Cultural Revolution in China or the reign of the Khmer Rouge in Cambodia), but also during periods of slow economic decline (e.g., as in much of sub-Saharan Africa during the 1980s). Although they may have had a secular tendency to move forward together during recent history, this relationship is historically contingent and not inevitable.

Each category of structural change can to a certain extent be seen as the result of applying new technologies to human efforts, where technology is defined broadly in “software” terms (e.g., accounting techniques or legislative procedures) as well as hardware terms. However, the types of technologies involved were specifically those developed in the West during the sixteenth through nineteenth centuries, and it does not follow that all types of new technology will lead to similar changes in interaction patterns.

77 Industrialization can be defined as the increased use of machinery in economic processes, commercialization as an increase in amount of products that are produced for trade rather than home consumption, and political centralization as an increase in the size of jurisdiction for authority structures.
Finally, it is obvious that the three broad categories of structural change do not comprise all possible sources of either broadening or fragmentation. What all this indicates is that it is probably impossible to come up with any general theory about the structural causes of either broadening or fragmentation of interaction patterns, much less to attribute them causally to some single variable such as “level of modernization.” However, one can still note that the types of structural changes (e.g., industrialization, commercialization, and bureaucratization) that occurred in the West up to the twentieth century (and continue to occur in the developing countries) tended to move interaction patterns toward increased breadth and fragmentation. This in turn justifies linking the theory of change presented here with existing theories of modernization.

**Breadth and Fragmentation of Interaction and Altruism Patterns**

The theory presented hence takes changes in patterns of interaction as its basic starting points. In this section, it is shown to provide a causal link between changes in patterns of interaction and an important cultural process often discussed in the literature on tradition and modernity: changes over time in patterns of altruism. Furthermore, it can shed some new light on aspects of the latter process that have long been a source of theoretical disagreement. Although classical social theorists were unanimous in the view that traditional societies were characterized by high levels of altruism that nearly fused identities within small communities, they were far from unanimous on the levels of altruism in modern societies. Tonnies viewed modernization as leading to isolated individualism, a “deaffectionation” of society, a theme that is also repeated in Parsons’s pattern variables. Durkheim, on the other hand, implied that a traditional type of altruism associated with primordial ties will be replaced by a broader, more modern type of altruism toward society as a whole. This latter implication has been picked up by scholars of “nation-building,” who argue that in the long run narrow, parochial loyalties will be replaced by broader ties to the nation-state. These disagreements are connected to larger questions about the sources of solidarity and integration in modern societies.

In the theory presented here, a link between patterns of interaction and altruism will be derived using the arguments introduced in chapter 3 and developed further in chapter 5. In these discussions, it was inferred from the coherence model that individuals engaging in cooperative interactions within particular group boundaries would develop positive altruism toward other members in the group. These discussions, however, concerned the causes of altruism within more or less static boundaries of collective interaction. In this chapter, analysis focuses on the effects of changes in the size of boundaries (broadening) and creation of subboundaries within a collective (fragmentation).

---

78 See discussions in Shils 1975b, 113.
81 Geertz 1963; Bendix 1964.
First, it can be shown that broadening of interaction patterns, ceteris paribus, leads to a broadening of the scope of altruism, without any diminution in its intensity. When the average size of cooperating groups increases, assuming no other changes in the pattern of interaction, neither the current marginal weighting benefit of the welfare of the group as a whole nor the optimal amount of adjustment in these preferences will change since neither depends on the size of the group (see theorem 3 of chap. 3, and discussion on altruism). Moreover, if an individual increases his or her weighting of the aggregate welfare of a particular cooperating group by a particular amount, this is arithmetically equivalent to an increase in the weighting of the personal welfare of each other individual within the cooperating group by the same amount. This indicates that a change in the size of the cooperating group will not affect the amount of change in altruistic preferences toward any individual within the group.

However, it can also be shown that increasing fragmentation of interaction patterns, ceteris paribus, leads to a decrease in the intensity of altruism. This can be seen by examining a series of $k$ cooperative actions by an individual in groups of size $m$, where the $k$ groups of $m$ are chosen from among a larger population of size $p$. This means, assuming that all member of the population are equally likely to be included within a particular cooperating group, that in the former case each other individual in the larger population will participate in an average of $km/p$ cooperating groups. Assuming that each interaction leads to an average increase of $Da$ in the valuation of the welfare of others within the cooperating group, this means that at the of the series of $k$ actions the individual will experience an average increase of $km/pDa$ in the average valuation of the welfare of others in the population. The most straightforward way of representing a decrease of the intersection between cooperating groups is by enlarging the larger population from which these groups are chosen from $p$ to $q > p$. Because $km/qDa < km/pDa$, the average increase in the valuation of the welfare of others in a population will be smaller as interaction patterns become more diffuse and complex.

Clearly, the relationship between interaction patterns and patterns of altruism is not unidirectional. Widening boundaries of altruism tend to create incentives for further cooperative interaction along those wider boundaries, while weakening the intensity of altruism weakens the incentives for cooperation altogether. However, it should be noted that given some long-standing and relatively static pattern of altruism and interaction the initial impetus for change must come from some exogenous factor. Since in the coherence model all sources of changes in altruism are generated endogenously, this means that the initial changes must be structural, based upon technological change or some other factor.

---

82 As noted, is equal to the sum of the individual welfares of group members, denominated in the goods that the individual’s actions are contributing to expected group outcomes.
The main, and somewhat counterintuitive, conclusion of this analysis for the relationship between “modernization” and the level of altruism in society is that the relationship is not simple or monotonic. There will in fact be two dimensions to the level of altruism, the breadth of the group toward whom the individual feel altruism, and the intensity (utility weighting) of that altruism. Nonetheless, it can be shown that in extreme cases of “traditionalism” and “modernity,” i.e., low broadening/low fragmentation and high broadening/high fragmentation, individuals will primarily be concerned with their own personal welfare, while high levels of concern for others will be found primarily in the intermediate stages.

The most extreme “traditional” case is the mythical “state of nature,” in which each individual cooperates within a group of one person (himself). This corresponds to the narrowest possible cooperation pattern, with no fragmentation. In this situation, the individual will clearly feel no altruism toward any other individual.

Of course, such a state of nature does not exist even in the most technologically underdeveloped society. In preindustrial, precommercial communities, intense patterns of cooperation occur within small, stable groups. Because of this, one can expect that individuals in such communities will feel very intense levels of altruism since the level of fragmentation is very low (see chap. 5). On the other hand, the breadth of that altruism will be very narrow and confined to the community.

As industrialization and commercialization proceed, the breadth of cooperation patterns will widen as individuals migrate into population centers and engage in factory work, work in the retail or service sector, or work in government. At the same time, the interactions will become increasingly fragmented as individuals begin to cooperate within different circles in their attempts to obtain different goods, hence decreasing the strength of altruistic ties.

At the other extreme of “modernity,” an individual will cooperate within a set of boundaries that encloses the entire population but patterns of cooperation will be highly compartmentalized and domain-specific, and the frequency of cooperative interaction with any particular set of individuals will be low. In this case, fragmentation will cause the level of altruism toward other individuals within the boundaries to be very small, although the boundaries will be very large.

Both of the extreme ends of this process of technological change can be considered as individualistic or “self-oriented” (to use Parsons’s terms) societies in that individuals will not be much concerned with the welfare of others in making their decisions. The unattached individual in the “state of nature” will be relatively free of altruistic ties to others, but so will the individual with the thinly stretched and crosscutting network of ties found in a technologically advanced society. What this theory predicts is that the most politically salient affective ties will be found primarily in the stages in between.

Patterns of Altruism and Collective Violence
Changes in patterns of altruism can be shown to have implications for a central behavioral issue in the modernization literature, the
234 Choosing an Identity

relationship between the structural and cultural changes associated with modernization and levels of political violence in society. This in turn is related to the more basic issue in political analysis that is often referred to as the “Hobbesian problem of order.”

Observation of politics in developing countries has brought up what Huntington calls the “rate/level paradox,” the perceived tendency for countries at high levels of socioeconomic modernity to experience less political violence than countries with low levels, while countries with high rates of movement toward modernity to experience more violence than countries with lower rates. An alternative, though somewhat similar, “paradox” was perceived tendency for countries at very low or very high levels of “modernity” to experience lower rates of political violence than countries between these two levels. Both of these types of tendencies were originally posited by a number of theorists of “transitional societies” and were later investigated by theorists of “breakdown” such as Eisenstadt and Huntington; economists such as Olson; theorists of revolution such as Davies, Gurr, and the Feierabends; and, indirectly, theorists of nationalism such as Gellner, Connor, and Anderson.

Within this literature, a wide variety of explanations has been offered for the increase in violence that is alleged to occur with the movement toward modernity. Some arguments focus on psychological/cultural factors such as anomie caused by the breakdown of traditional identities, heightened expectations caused by economic growth, dissatisfaction created by uneven development and/or heightened political awareness due to education and the mass media. Others focus on structural factors such as the increased incentives for individuals to organize into large competing groups and the resources to do so created by improved communications and transportation. Others focus primarily at the behavioral level on increases in political participation and the spillover of this into violence and instability. On the negative side, increasing levels of violence are attributed to the relative weaknesses of political institutions and their inability to provide

---

83 Huntington 1987, 17.
84 Olson 1963.
85 Davies 1962; Davies 1969; Gurr 1968; 1970; Feierabend, Feierabend, and Nesvold 1969. See also special issue on Revolution and Social Change of Comparative Politics 5, 3 (April, 1973).
87 Pye 1962, chap. 2.
88 Davies 1962.
89 Gellner 1969, chap. 7; Olson 1963.
90 Lerner 1958; Deutsch 1963; Gurr 1970, chap. 2.
91 Gellner 1969.
sufficient authority and integration to overcome societal propensities toward violence.\footnote{Deutsch 1963; Eisenstadt 1973a; Huntington 1986}

Investigation of the phenomenon became more intense following Huntington’s publication of Political Order in Changing Societies, both because of the prominence of the book and because it clearly laid out the case for this view of political violence.\footnote{See pp. 39-59.} This brought forth a number of critiques and explorations of the issue of modernization and violence, some of which questioned whether the empirical tendency being analyzed really exists.\footnote{Tilly 1973; Ake 1986.}

On the whole, although investigation of these issues has often been quite detailed, it has not served to justify the conclusion or put forward an alternative view of the relationship between social change and violence. One reason, of course, may be difficulties in operationalizing and measuring levels of violence. A deeper problem lies, however, in the ad hoc nature of their propositions about the effect of socioeconomic changes on political action as well as the lack of systematic analysis about the intermediary variables of preferences and beliefs. As with modernization theories in general, the lack of a clear set of assumptions underlying propositions about the influence of structure on behavioral change promotes theoretical incoherence.

Furthermore, because existing theories lack such microanalysis, they fail to provide clear logical links in their analysis of structural and cultural change or of their analysis of cultural and behavioral change. Why should rapid industrialization, commercialization, and bureaucratization lead to heightened anomie and frustration? Why should heightened anomie and frustration lead to increased political participation, including violence? Theorists who examine the subject generally take the causal links as ad hoc assumptions rather than generating explanations for them.

This lack of explanation is particularly problematic given the history of conventional rational choice attempts to analyze individual participation in large-scale collective political action. As was noted in chapter 1, plausible structural assumptions and conventional actor assumptions seem to indicate that the expected marginal personal benefits of participation in large-scale action will rarely exceed the marginal costs, regardless of an individual’s interest in the success of the action, and individuals will generally be predicted to free ride.\footnote{Olson’s work on modernization predates his work on collective action, and implicitly assumes individual participation in group action. It is also noteworthy that this pioneering effort by a contemporary economist to explain long-term political change eschews conventional rational choice analysis, bringing in such factors as the “frustrating severance of social ties” in Olson 1963, 535; and the “revolution of rising expectations” in Olson 1963, 541 as well as “charismatic leadership, religious controversy, (and) ideological change” in Olson 1963, 543.} This is particularly true for participation in collective violence, given the high level of personal risks involved. However much cultural or structural
changes may give certain groups of individuals a collective incentive to violently oppose the status quo, the risks of violence to personal welfare seem to make individual participation irrational\footnote{Barry 1970, 44-46; Salert 1976, chap. II, 33-38; Buchanan 1979; T. Mason 1984, 1041-43; Elster 1993, 15-17.} regardless of levels of heightened political awareness or discontent. Hence, the very existence of political violence is seen as something anomalous, no matter how socioeconomic change is proceeding.

However, as was discussed in chapters 3 and 5, the existence of altruistic preferences toward a group can be a force that makes it rational for individuals to participate in mass political activities, despite expected personal costs. This is particularly true for violence; the importance of altruistic identification for individual participation in collective violence is an issue that has been noted by various theorists from Machiavelli on.\footnote{Machiavelli 1950, chap. 12; Shils and Janowitz 1948; Coser 1956, chap. 6; Connor 1969; Muller and Opp 1986.} By compensating for personal risks through benefits for one’s group, altruism can make participation in even very risky activities rational. Furthermore, altruism toward one group, in this view, can become the instigator of violence toward another.

Even if one accepts the significance of altruism in collective violence, however, a nontautological theory cannot simply assume the existence of altruism in order to explain participation; it must explain the sources of the altruism in a systematic manner and account for its intensity and extent. Furthermore, it must show how altruism affects decisions to participate.

In the formulation presented here, the level of violence in society directed against the state is related to the resources for coercion available to groups who may oppose or seek to bypass the constituted authority of the state relative to the resources for coercion available to it. As long as the former far exceeds the latter, the state will be able to deter violent opposition since it will be irrational for groups to attempt to violently challenge state power. Hence, at least in the long run, the level of violence in society will be relatively low.\footnote{This of course does not cover the possibility that the state itself will use violence against groups in society. However, assuming that the state is rational and has sufficient resources to subdue opposition, it will eventually be able to sustain the level of obedience it desires without the use of violence.} I do not attempt to account for the level of constituted authority in a particular society, a complex issue that is beyond the scope of this theory and remains unresolved by rational choice theorists.\footnote{See discussions in Hardin 1990; 1991.} Instead, for the most part I try to predict the resources available to groups that are potentially opposed to the state, which in turn will relate to the expected level of violence in a society given a particular level of constituted authority.

The resources available to a group will be related both to the size of the group and the probability that they will act cooperatively in...
violent action against constituted authority. These factors are clearly related to the patterns of altruism in society since the probability of cooperation will be closely related to the average intensity that members of the group feel toward one another relative to the intensity of altruism that they feel toward other members of society since altruism reduces collective action problems (see chap. 3). Finally, using the analysis of the previous section, levels of altruism can be linked back to patterns of earlier cooperative economic and political interaction and the coherence-linked preferences they generate.

With this point in mind, it is possible to investigate the potential for violence in societies with different patterns of interaction. Three categories are put forward, each of which is associated with different patterns of altruism, which in turn are promoted by a different pattern of human interaction on the “modernization” scale.

**Primordial Society**

The most “traditional” pattern of altruism is that in which interaction takes place within very narrow boundaries. Given the small size of interacting collectivities, it is unlikely that interactions will be very specialized and low levels of fragmentation can be taken for granted. Such a pattern of interaction can be said to exist among hunter-gatherers, nomadic pastoralists and subsistence farmers. Such patterns of interaction will lead to patterns of altruism that are very intense but directed at a relatively small group, a pattern that can be called *primordial* altruism. At its very narrowest, when cooperative activities are confined within a relatively self-sufficient family unit, altruism will be very narrowly directed in a way that is extremely parochial, corresponding to the “amoral familism” discussed by Banfield and others.

This type of society, without a state, resembles the Hobbesian “state of nature,” which is enveloped by the violence of individuals or atomized groups fighting against one another, the “war of every man against every man.” However, it is also a society in which authority is relatively easy constitute because of the atomized nature of the individuals or groups that might oppose the state. A relatively low level of authority can effectively deter opposition. Because of the narrowness of their altruism, individuals will not be able to unite into large groups without suffering from collective action problems that leave them incoherent in joint action against the state. The difficulty of joint action against the state will hence leave individuals largely apathetic in political matters, a state of affairs that is in many ways conducive to societal stability.

This does not mean that violence will not occur in such societies, but violence will take place at a very low

102 Hence, for simplification purposes, the resources available to different resources will be seen to be equal and the combined resources of a group will be seen as equal to the sum of resources of members, provided that they all act cooperatively.

103 Banfield 1958.


105 See discussion in Rule 1988, 146-47.
level between small groups of individuals and in a way that does not threaten the state.

Plural Society
In societies in which interaction patterns are broadening but fragmentation has not taken place to a great extent, altruism will tend to be directed at a larger group than in a primordial society, but will still be relatively intense. Such a society can be called a “plural society” after the terminology of Furnivall and Smith. It can be found where industrialization and commercialization have proceeded to the extent that individuals are coming into broader patterns of interaction within population centers, cementing patterns of cooperation that fall beyond the narrow bounds of rural communities.

Such societies are those most frequently associated with the alleged increase in violence that supposedly accompanies modernizing structural changes. As has been noted, technologically driven socioeconomic changes has often been thought to lead to increases in levels of violence. Furthermore, as was noted in chapter 5, much of this increasing violence is viewed to be broadly “ethnic” in nature.

A microlevel explanation for this can be found in the fact that broadened interactions promote the formation of broader group identities. When individuals are migrating from rural communities of origin to urban population centers the most efficient boundary for such new identities is generally one based on some set of ascriptive criteria, resulting in large-scale ethnic groups. Furthermore, as such groups become entrenched, altruism is transferred from communities of origin to the ethnic groups as a whole.

The formation of large-scale ethnic groups in turn allows for large-group actions that may overwhelm the enforcing capabilities of constituted authority. In such a situation, order can only be maintained if the state is supported by at least one of the large groups. Moreover, given the division of altruism patterns into large groups, a coalition of more than one group to control the state will always be subject to internal conflict. Hence, it follows that order can be maintained only if one large group has a monopoly on state power, a conclusion that conforms with the analysis presented in the literature on plural societies. It also conforms with the analysis of those who argue that the existence of large-scale ethnic groups is antithetical to the preservation of a stable democracy.

The only way to prevent this situation is if fragmentation takes place to a sufficient extent to weaken the strength of altruism that an individual feels toward any particular ethnic group. In this case, the broadened boundaries of altruism will not be sufficiently strong to support cooperation within these group boundaries, including violence against other groups. This weakening of altruism due to

---

106 Furnivall 1939 and Furnivall 1948; Smith 1965.
108 Smith 1965, 86.
109 Lipset 1959a, 31.
the specialization and compartmentalization of interaction patterns corresponds in large part with the phenomenon of “crosscutting cleavages” that has long been seen as a guarantor of stability in plural societies.  

Civil or Mass Society
What this indicates is that broadening of interaction patterns, at least initially, tends to increase the levels of violence that can be expected in a society unless they are offset by sufficient accompanying levels of fragmentation. However, at a certain point the relationship and broadening begins to move in the opposite direction. Specifically, when broadening becomes wide enough to promote altruism patterns that encompass the bulk of a society, the pattern of altruism becomes transformed from one that is divided into large groups that are internally altruistic into one in which society itself begins to constitute a single group. At this point, an authority can be constituted which commands the support of a vast majority of the population against any groups that would seek to challenge it.

Here, we can see the beginnings of what is usually called “civil society,” a phenomenon requiring that all individuals in a society have at least a modicum of positive regard for all other members. Such societal integration was seen by Parsons and his followers as the only possible solution to the “Hobbesian” problem, and it was argued that societies, in order to be stable, require some set of shared values and could not be held together by pure self-interest. This idea has been developed by modernization theorists, who have argued that a lack of integrated values is the greatest threat to stability in modernizing societies.

However, fragmentation may be proceeding apace with broadening, in which case a pattern of altruism that encompasses an entire society may be very weak. This situation corresponds to what is usually referred to as “mass society,” a society exhibiting both the structural attributes associated with modernization and a highly atomistic political culture. In this case, the “interpersonal bonds of common locality, ethnicity, religion and class are severed,” but no set of civil or national-level ties are created to replace them. Ironically, such a society is much like the most “traditional” society in that it is unlikely that large groups will form to oppose state authority since few individuals would be willing to undergo personal risk to protect the state. Because of this, only a relatively low level of constituted authority will be necessary to deter opposition to the state.

---

110 Original statement in Bentley 1908. For more recent discussions, see Coser 1956, 76-79; Lipset 1959a, chap. 3; Parsons 1969.
112 Parsons 1951, 36-37.
113 Weiner 1965; Binder 1964.
114 Kornhauser 1959.
115 Sztompka, 80.
Order versus Democracy
A number of points can be deduced from the preceding analysis: The first is that different types of societies present different types of solutions to the “problem of order.” In a primordial society, order can be maintained by a relatively weak state authority because of the narrowly focused patterns of altruism in society. In a relatively unfragmented plural society, order can only be maintained, with some difficulty, by the dominance of a single ethnic group. In a relatively fragmented plural society or a fragmented mass society, order can again be maintained by a relatively weak state authority due to the weakness of altruism patterns vis-à-vis concern with personal welfare. In a relatively unfragmented civil society, order can be maintained by the mutual concern that individuals have toward other members of society and a state that can hence command a certain amount of consensual support.

Another point that follows from this is that if both broadening and fragmentation are seen as characteristics of modernity then the most “modern” society is in some ways the closer to the most “traditional” than it is to the others along the continuum. In both of these types of societies, little altruism is felt toward larger groups. In the “traditional” (primordial) case, this is because altruism is directed intensely toward narrow groupings. In the “modern” (mass society) case, it is because altruism is very weak overall and personal self-interest dominates. In either case, a relatively low level of constituted state authority will be sufficient to maintain order.

This particular type of order, however, is not likely to be democratic, because state control can be dominated by a relatively small coherent group, which can maintain power due to the atomized nature of possible oppositions. In a primordial society, this dominating group will be a small, highly coherent elite. In a fragmented plural society, it will be the ethnically and economically bounded subgroup that best manages to avoid internal fragmentation. In a mass society, it will be a group whose interests are most closely tied together by economic characteristics. Fragmentation prevents communal polarization, but it also allows a relatively weak but coherent group to take power without organized opposition. Hence, it can promote authoritarianism. What this further indicates is that democracy, if it is defined in the most stringent sense, can only be maintained within a relatively unfragmented civil society.

It also means that the conflicts between goals of order and democratization create different sorts of tradeoffs at different stages in the broadening and fragmentation process. Fragmentation is always good for minimizing levels of civil violence, although the order that is maintained is not likely to be very democratic. It reduces the incentive for individuals to engage in collective violent action, yet at the same time lowers the obstacles to a small, highly solidaristic group seizing and maintaining power. Broadening initially increases levels of civil violence by increasing the incentives for political violence and hence decreasing the effectiveness of a particular level of constituted authority, but eventually it creates the conditions necessary for a democratic civil
society through a final broadening that includes all members of a society within the boundaries of each individual’s altruistic preferences.

Hence, if one assumes that a democratic order is preferable to authoritarian order but that any kind of order is preferable to no order at all, the problem of development is even more complex than that sketched by Huntington.116 Broadening is initially bad, then eventually good, while fragmentation is initially good, then bad. The problem is that both of these processes tend to move in a single direction and it is difficult to reverse either of them.

Conclusion

In summary, the theory presented in this chapter makes an attempt, albeit a sketchy one at this point, to explain at a microlevel some ways in which structural change can lead to changes in culture, and how this can lead to changes in behavior. Furthermore, it does so through a parsimonious set of assumptions included in a broader model of individual action. In doing so, it shows how the influence of structural change on culture is not amorphous or ad hoc but can operate systematically through two processes, broadening and fragmentation. Finally, it shows how the different kinds of altruism patterns caused by these processes lead systematically to different patterns of violence and different solutions to the “problem of order.” In doing so, it provides a system of simple causal linkages between a jumble of phenomena that have previously been presented as macrocharacteristics of modernity.

Chapter 7

Conclusion

In concluding, I will try to present ways in which the theory can be augmented and improved as well as justifications for its basic shape and form. As was stated in the beginning of this book, I do not mean to propose that there is a single model which can predict all of human behavior or that a single actor model should be hegemonic in social science analysis. I simply attempted to provide a deductive model of preference and belief formation and to examine its implications in conjunction with the rational optimization assumption for empirical problems in comparative politics. Given this, it is reasonable to see the model presented here as one that is subject to considerable improvement, and to consider some of the ways in which this may be done. At the same time, it is important to discuss why its shortcomings are a cause for refinement rather than abandonment.

Modifying the Assumptions of the Model

The changes discussed here are based upon modifications rather than replacement of the rational optimization portion or the coherence portion of the model. Although it is conceivable that alternative decision-making models to rational optimization could be devised, existing alternatives either fail to yield assertive predictions or are based upon modifications of or qualifications (bounds) on rational optimization. Furthermore, as was noted in chapter 2, rationality of some sort corresponds to the way in which individuals intuitively make sense of each other’s behavior. The major theoretical competitor to rationality, action based upon norms and procedures, has yet to produce an assertive general model. Moreover, given that most versions of such models imply that norms and procedures are absorbed passively by actors from their surrounding environments, an assertive general model of nonrational action will require an assertive general model of structure. As noted, it will be quite difficult to come up with such a model.

The alternative preference and belief formation models to coherence-seeking are less clear, but in order to maintain generality and assertiveness while avoiding the problem of having to determine environment-specific parameters any alternative will have to be based upon preference and belief adjustment that maximizes or minimizes some quantity. As was pointed out earlier, expected utility cannot be this quantity since it is always possible to adjust preferences in order to provide arbitrarily high amounts of expected utility. Furthermore, one must also choose a quantity that does not cause beliefs to violate the information. This rules out quantities that can be determined through direct observation, such as one’s wealth, fame, or physical strength. Likewise, highly abstract normative quantities such as goodness can be ruled out because the impact of particular preference and belief changes on the such quantities are unclear and the resulting model will
be unassertive. What remains is some sort of relationship between preferences and beliefs and past actions, and it can be said in support of the model presented here that building a coherent identity out of past actions, preferences and beliefs is perhaps the most straightforward. Furthermore, it corresponds quite nicely with the common notion that individuals are constantly being remade and transformed through the process of action.

Given this, three possible to the model in particular are discussed here: two have to do with bounded rationality: nonexpected utility and satisficing. Another has to do with norms: focal points. Finally, there is nonrationality.

Nonexpected Utility
One major possible addition to the model is the incorporation of framing effects, i.e., effects that transform preferences and/or beliefs prior to decision-making process for a particular choice set, a transformation that remains in effect only for that particular choice set. Framing effects operate in two different ways. Those found in the value portion of prospect theory as well as regret and disappointment theories, describe the effect of frames in transforming preferences. On the other hand, those found in the decision weights portion of prospect theory and other nonexpected utility theories, describe the effect of frames in transforming beliefs. The first type of transformation is compatible with the rational optimization assumption, while the second may not be.

The transformation of preferences by framing comes in two main known varieties: endowment effects, in which preferences are transformed as a function of the amount possessed or consumed of desired or undesired goods and choice effects, in which preferences are transformed as a function of the various alternatives available in a choice set. Endowment effects can be found in the value portion of prospect theory as well as in certain types of addiction and satiation theories. Choice effects can be found in regret and disappointment theories. Choice effects can be found in regret and disappointment theories (see chap. 2).

Although the volatility of preferences as a result of framing stretches the notion of what it means to be “rational,” all of these effects are compatible with the rational optimization assumption, which simply requires that preferences be well ordered and defined over all possible outcomes. Furthermore since they act to transform some existence set of preferences, they depend on some model that can explain the untransformed preferences. One such model is the coherence model presented here. Furthermore since framing effects on preferences are transitory, they do not contradict any predictions that such a model will make about untransformed preferences in future periods.

On the other hand, all preference framing effects models include some sort of unspecified parameter such as the steepness of risk-seeking and risk-aversion curves around the status quo for the value portion of prospect theory or the extent of risk aversion with regards to regret in regret theory. Such parameters must be determined in some manner if the resulting integrated model is to be assertive.
Framing effects that transform beliefs are found in nonexpected utility theory, and typically function by transforming the perceived probabilities attributed to particular outcomes. For instance, prospect theory assumes that individuals base their decisions on transformations of perceived probabilities that increase the relative weighting of very likely and unlikely outcomes. Such effects may contradict the rational optimization assumption since the transformed probabilities may contradict available information in those cases in which objective probabilities are known. One can still use such effects in conjunction with the rational optimization assumption and some model that can account for untransformed perceived probabilities; they can be seen as a kind of “friction” on perfect optimization. On the other hand, each of these effects also has unspecified parameters concerning the extent of transformation that needs to be determined to ensure assertiveness in the integrated model.

Given the strong empirical evidence that certain framing effects do occur in experimental settings and the fact that they are compatible with the coherence model, it seems likely that attempts to refine the model ought to include recognition of such effects. Since so many different framing effects have been put forward, however, it is not clear which ones ought to be adopted.

Focal Points
Focal points are another way in which beliefs may be transformed at the time of choice. The idea of focal points was originally put forth by Schelling in his work on military conflict, and it has recently been extended to analysis in a number of other areas, including corporate culture and international political economy. Focal points become relevant in situations of strategic interaction between actors, where each actor’s existing beliefs are insufficient for predicting the actions of other actors, which in turn makes it impossible for him or her to choose an optimal action. Under such conditions, certain actions may be chosen simply because they are highlighted in some sort of way by the environment and will be optimal if other actors react similarly.

In this case, the transformed beliefs will have to do specifically with expectations about the actions of other actors, and will be generated only when existing beliefs are insufficient to make assertive predictions about such actions. Because of this, beliefs based upon focal points do not contradict the assumptions of the coherence model since they are created concerning the actions of a particular actor only in the absence of coherence or information-generated beliefs about such actions. As Schelling’s original formulation put it, actors will not search for focal points as predictors of others’ behaviors if they already have sufficient beliefs to make such predictions.

1 Schelling 1960.
2 Kreps 1990.
3 Garrett and Weingast 1993.
4 Schelling 1960, chap. 3.
5 Schelling 1960, chap. 3.
Because of this, focal points may add an element of additional assertiveness when the coherence model cannot predict beliefs. Unlike framing effects, focal points add probabilistic beliefs about outcomes rather than simply transforming existing beliefs. On the other hand, however, focal points do not currently form an assertive model. There currently exists no general set of assumptions that can predict which types of actions will be chosen as focal points within a particular choice set; this depends on the specifics of the environment in which the choice takes place.

A set of somewhat assertive assumptions can be generated, however, by noting that historical precedents often form the basis for focal points. Actors in general will assume that other actors will behave as they have in the past, as long as such behavior is not clearly irrational. This is captured somewhat already in the idea of a Nash equilibrium, in which actors in long-term iterated interactions are assumed to believe that other actors will continue the same patterns of behavior that they have conducted in the past as long as it is not clearly suboptimal for them to do so. The nature of the equilibrium is obvious, however, only when characterizing past patterns of behavior is unproblematic. When this is not so, the broader notion of a focal point is necessary to make predictions about behavior, and a certain amount of context-specific interpretation is necessary to determine which actions will be highlighted.

**Time preferences**

When the effects of actions are spread out over several periods, another elaboration that can be made is time preferences, i.e., nongeometric discounting of future utility. As was noted in chapter 2, there are many different types of nongeometric discounting, and an assertive model that incorporates a designated set of time preferences needs to specify the functional form of the discounting as well as any additional parameters entailed.

However, the coherence model views time preferences in a different manner. The offshoot of time preferences that applies specifically to the coherence model is the idea that individuals may actually be able to manipulate their utility discounting functions in order to minimize coherence. Although this has not been discussed explicitly, there is nothing in the assumptions of the coherence model that disallows adjustment of utility discounting, so it requires no real modification of the model. Furthermore, the model allows for the possibility that individuals may have different rates of discounting for different goods, a kind of change that generates various self-management issues, even when each of the discounting functions is geometrical.

Framing effects, focal points, and time preferences are not the only modifications that can be made. Another is the imposition of the “hyperrationality” assumption that actors will always adopt preferences and beliefs that isolate a single action within a choice set as optimal; this implies that individuals will make such adjustments before as well as after actions. An even more major modification would be the possibility that actors might actually engage in “deliberate character planning,” attempts to affect their future preferences and beliefs by
deliberately putting themselves in situations in doubt-reduction leads to certain types of preference and belief change.\textsuperscript{6}

Such elaborations, while potentially very interesting, introduce great new indeterminacies into the model, and their empirical payoff is unclear. The imposition of hyperrationality does not provide any indication of the types of preferences and beliefs that will be adopted to allow clear optimal choices for every choice set. The introduction of deliberate character planning does not specify the criterion by which individuals will attempt to influence their future characters, or how this trades off against current attempts to maximize expected utility and minimize cumulative doubt. Finally, the additional leverage that they provide in explaining empirical social, political and economic phenomena is unclear.

On the other hand, framing effects, focal points and time preference are modifications to the model that are addressed at specific empirical inaccuracies, and at least have the potential to lead to more assertive and accurate predictions. Incorporation of framing effects and time preferences can remedy potential empirical inaccuracies of the model with regard to the weighting of different choices in a choice set and with regard to the distribution of utility over time, while the addition of focal points will improve the assertiveness of the model by partially isolating some additional sources of beliefs where choice would otherwise be unassertive.

\textbf{Justifications for the Coherence Model}

In order to view this model as the appropriate base upon which to build additions rather than as a mere curiosity, it is useful to examine and attempt to justify some of the basic characteristics. The main characteristics examined include the fact that it is an individual-level actor model, that it seeks to incorporate cultural variables into such a model, and that it places relatively high level emphasis on generality and parsimony.

\textbf{Why an Individual-Level Actor Model?}

One justification for the model presented here can be based on a version of methodological individualism. One should of course be wary of using this term because of the multitude of definitions that have been attached to it. There are really two different major types of methodological individualism: The first argues that the individual level is the only appropriate level for analyzing human phenomena.\textsuperscript{7} The second suggests that individual-level analysis is but one of a number of different appropriate levels at which one may studying human phenomena.

My own philosophy falls somewhere between the strong and weak types. I do not believe that social structure and culture are

\textsuperscript{6} These types of effects are discussed in the introduction to Elster 1993 as well as Elster 1983d, chap. 3.

\textsuperscript{7} This is roughly Elster’s position, that individual action should be seen as the “basic building block of aggregate social phenomena.” See Elster 1993, 8.
unimportant causal factors, nor that higher-level analysis is not useful, or that all higher-level phenomena can readily be explained by means of individual-level theories. However, I believe that it is important to attempt to eventually explain all known higher-level outcomes as products of individual-level actions.

One reason for this is that all social phenomena, whether intended or unintended, are causally the result of individual actions. Such a position thus accepts the famous Homans quote, “If you look long enough for the secret of society you will find it in plain sight: the secret of society is that it was made by men, and there is nothing in society but what men put there.” However, it does not draw from this the typical conclusion that only individuals should be studied. It is certainly plausible to argue that society as a whole is a construction of individuals, yet argue also that that it is important to study structural as well as individual phenomena. After all, there is nothing in men but what cells put there, yet it would not follow that all analysis of human beings should start at the cellular level.

Such a position also accepts the sociological notion of emergence, i.e., the idea that the basic characteristics of a system cannot be reduced to the relatively simple patterns in the behavior of its lower-level constituents. This notion has been adapted more generally by theorists of complexity with regard to all types of complex systems, as well the relatively simple units that make them up. Indeed, the example of complexity theory is important since work in this field shows that translation from a lower level to a higher one another can lead to emergence, but also to a kind of “deemergence,” where orderly action on the individual level can look chaotic when taken to a higher level.

Hence, it is important not to view theories at differing levels of analysis as competitive, but as ultimately complementary. Furthermore, it is inappropriate to ignore clear patterns in human existence simply because they are at the “wrong” level of analysis. There will always be emergent phenomena at higher levels of analysis that cannot be explained straightforwardly in terms of phenomena at a lower level of analysis. Nonetheless, it is important for theorists working at one level to try, if possible, to show what their models would predict for patterns at both higher and lower levels of analysis and hence to

---

8 See Homans 1961, 385. A more extreme version of this viewpoint is Thatcher’s alleged statement that she couldn’t comment on the state of British society because such an entity did not exist; only British people existed.

9 The term was originally coined by Parsons in his discussion of social systems and individuals. See Parsons 1937. Parsons based his idea upon his analysis of Durkheim’s critique of utilitarianism. See for instance Durkheim [1893] 1984, 220-21.


11 This is typified by work in this field on artificial life and societies. See Levy 1992.
draw connections between their models and those models at these other levels.\footnote{For an interesting discussion of the need to simultaneously explain phenomena at several levels of analysis, see Jerome H. Barkow, Levy 1978; 1980.}

Another justification of this limited methodological individualism is that our everyday perceptions are of individual-level actions not of society as a whole, and it is thus reasonable to expect that social science theories can be stated in terms that are amenable to such everyday confirmation. There is some sense in which theories are often easier to make sense of when they are stated at the individual level than when they are stated at the structural level. The typical person will be less concerned with macrovariables such as social class structure than with his or her life and the life of other individuals. Of course, it is usually the case that such macro-variables have a very strong relationship to an individual’s own circumstances, but in order for this to be clear the individual-level causes and consequences of structure have to be known. This in turn necessitates the translation between levels that I have just mentioned.

This justification also relates to the comparison within methodological individualism between intentionalist, behaviorist, and connectionist theories. In addition to the defense of preferences and beliefs mentioned in chapter 2, it is important to note that preferences and beliefs are not constructs invented by social scientists. Rather, they are an essential part of the way in which human beings ordinarily make sense of one another’s behavior in everyday life, as is taken for granted in the entire debate over “folk psychology.” Everyday attribution is usually in terms of motives (“What was that person trying to accomplish?”) and cognitions (“How did she believe she can go about accomplishing what she wants?”) rather than abstract rules connecting circumstances to behavior, much less circumstances to neural states to behavior. In a practical sense, the ability to use a theory to generate predictions relies on the ability of the implementor not only to select the theory that is appropriate for a particular empirical setting, but to quickly recognize the local phenomena that correspond to the variables described by the theory she selects. This in turn is difficult to do unless the theory fits in some basic fashion the way in which he or she intuitively makes sense of his or her setting. It is irrelevant to argue that such an implementor ought to develop a different mode of intuitive thinking unless there is some practical and reasonably uncoercive process by which this can be accomplished.

Why Integrate Cultural and Rational Choice Models?
One of the major features that distinguishes this model from the conventional rational choice model is that it attempts to introduce culture into rational optimization. This brings up a number of questions about the need for cultural variables or rationalistic assumptions, and about the overall usefulness of integration.

Until recently, culture and rational choice have been seen as opposite ends of the theoretical spectrum, as alternatives rather than
Conclusion as the basis for a combined model. Social science theorists have often drawn a contrast between “economic” and “sociological” approaches to analyzing human behavior, in which the economic approach is defined as conventional rational choice and the sociological approach is (more vaguely) defined as being analysis that uses collectively held preferences and beliefs, i.e., culture, as its primary explanatory variable.

This comparison has spawned two famous phrases in social science, the first being Duesenberry’s observation that “…economics is all about how people make choices. Sociology is all about why they don’t have any choices to make.” The second is Wrong’s characterization of sociological man as “oversocialized” and economic man as “undersocialized.” Political science has acted as arbiter between the two approaches, though hardly a neutral one, with its enthusiasm having swung from the sociological to the economic approach within the period from the 1960s to the 1980s.

The labels are currently somewhat misleading in that much of sociological theory today is avowedly “structuralist,” which is usually taken to imply that variations in intrinsic preferences and beliefs are relatively unimportant in predicting action. On the other hand, conventional rational choice still encompasses nearly the entire discipline of theoretical economics, despite the increasing attempts of economic theorists to incorporate cultural variables into their analysis.

In addition to the claim discussed in earlier chapters that the cultural approach is difficult to falsify, additional criticisms can be placed within two main categories. The first is that such an approach cannot account for the effect of structure on actions. Critics of cultural theorists have long accused them of assuming a simple causal link between culture and behavior and of denying the impact of formal institutions and other structural factors on behavior. This criticism is overstated, given the frequent links of cultural analysis in sociology and political science with systems analysis (particularly structural-functionalism) and its corresponding concern with the interaction and “fit” between culture and structure.

---

14 Barry 1970, 175.
16 Another change in sociology, one that reflects influences from cultural studies and anthropology, is the tendency to redefine culture as “physical culture,” e.g., artifacts and patterns of action, rather than as attitudinal. This essentially merges culture into structure and/or action, and makes such an approach essentially “structuralist” as well, despite the label, hence its frequent juxtaposition with some of the essentially materialistic general structural assumptions mentioned at the beginning of this chapter.
17 See discussions in Barry 1970, 4-6 and Lea, Webley, and Young 1992, 3.
18 Mitchell 1969, 103; Rogowski 1974, 4-11; Bates 1990, 31-35.
Furthermore, a number of theories within the cultural approach do explicitly take into account the effects of structure on action.\textsuperscript{20} However, it is true that the cultural approach as whole does not contain a general and assertive set of assumptions about how cultural and structural factors interact to determine behavior. While cultural analysis is not inherently incompatible with structural analysis, its lack of a general deductive model of how structure affects behavior often hampers its ability to make falsifiable predictions. This in turn suggests that such shortcomings can be remedied by integrating cultural analysis with the rational optimization assumption, which in turn means that the resulting integrated model need not generate the same difficulties.

A second set of criticisms asserts that the cultural approach focuses on how culture can affect action and structure but not on how action and structure can affect culture.\textsuperscript{21} As with the first set of criticisms, however, these are overstated when applied to the approach as a whole. The danger of ignoring the effect of structure on culture was recognized quite early among cultural theorists themselves.\textsuperscript{22} Furthermore, the cultural approach has produced a number of theories that analyze how structural changes affect culture in the medium term,\textsuperscript{23} as well as one major body of literature, modernization theory, that often focuses on the long-term relationship between structure and culture.\textsuperscript{24}

Nonetheless, it is true that the approach lacks a general, deductive model of how action and structure affect the preferences and beliefs that comprise culture. For this criticism, however, it is clear that integration with a rational choice model in and of itself is not sufficient. After all, the “answer” provided by conventional rational choice is far from satisfactory. It simply posits that the environment can provide information that affects individual and collective beliefs, and that it has no effect on underlying preferences, which effectively reduces culture to information. Hence, the need is for a model that not only is consistent with cultural factors and with rational choice, but also attempts to predict culture.

These two criticisms relate to a broader set of frequent criticisms which state that the cultural approach cannot explain conflict or change. The reason for this particular criticism seems to lie with the approach’s frequent association with structural functionalism, along with the tendency by some theorists to use the structural functional framework teleologically (see discussion in chap. 2, in section on structural methodologies). Since neither the cultural approach nor structural functionalism contains assertive assumptions, there is


\textsuperscript{21} Barry 1970, 48, 51; Pateman 1971; 1980.

\textsuperscript{22} Pye 1968, 224. See also Almond 1990c, 144.

\textsuperscript{23} For an exploration of the structural-functional approach to cultural change, see Lipset 1975. For a more recent formulation, see Eckstein 1988.

\textsuperscript{24} A discussion of modernization theory will follow in chapter 6.
nothing in either that prevents theories of conflict and change; several such theories are listed in the citations in this book. The real issue, again, is whether a general and assertive set of assumptions can be found that uses culture as an explanatory variable and can account for conflict and change.

As has been argued throughout this book, a general deductive model can be created which takes into account the sensitivity to variations in preferences and beliefs between actors, as well changes within actors over time. Predicting these variations requires incorporating some of the insights of psychological findings into rational choice theory, but I hope I have shown that this can be done in a way that preserves the elements of the theory that make it a useful tool for analysis.

Why a Parsimonious and General Model?
Finally, it is useful to take another look at the basic criteria presented in chapters 1 and 2 to analyze the conventional rational choice model and other models. I hope that the intervening chapters have shown that the coherence rational choice model presented here provides greater accuracy than the conventional rational choice model, yet retains a level of assertiveness that approaches or occasionally exceeds that of the conventional model. Furthermore, it is clear a priori that it retains the generality and parsimony of the conventional model, which in turn is the main parsimonious general model of individual behavior in the social sciences.

While the usefulness of assertiveness and accuracy for theories is rarely questioned, the same cannot be said for generality and parsimony. Hence the focus in this section is on developing the justification for these characteristics, building on discussions in the earlier chapters. While it is impossible to go into much greater detail here, it is possible to elaborate on them.

Theories that are designed to explain only one or a few cases are often difficult to use for predictive purposes since it is unlikely that their scope conditions will be met in the future. This also means, however, that it is difficult to falsify such theories and show whether they have provided anything deeper than a nuanced observation of past events. It also means that the policy implications of such theories are often unclear since they refer to circumstances that are different from those described by the theory. Nor is the solution to rely on a large collection of such theories since future circumstances will often fall outside the scope of any established by small groups of precedents.

Under rapidly changing conditions, therefore, theories based on a general model will often be the only ones that can generate predictions of any kind. As such, they can be seen as useful even if their predictions sometimes prove inaccurate. One irony of this position is that optimizing rational choice theories are, from the theorist’s point of view, a heuristic for estimating outcomes when detailed information about the preferences, beliefs, and decision-making processes of actors involved are unavailable. Most theories that assume that actors use heuristics, on the other hand, are usually designed to generate optimal explanations, i.e., to account for as many nuances of action as possible.
However, in doing so they make use of actor-specific information which goes beyond the level of what the typical applied researcher or policymaker can supply, and are unassertive without such information.

The argument for valuing relative parsimony is equally strong when one considers the long-held dictum in the philosophy of science against the use of “immunization” strategies to protect a general theory. As Popper noted long ago, “It is always possible to find some way of evading falsification, for example by introducing ad hoc an auxiliary hypothesis or by changing ad hoc a definition.”\(^{25}\) Allowing the unlimited addition of complexifying ad hoc assumptions is no better than an excessively narrow scope in keeping a theory useful for predictive purposes. The former will allow additions to a theory that exist only to accommodate one or a few cases, rather than integrating results extending over a variety of cases. There is little reason to have confidence that such additions will improve the accuracy of the theory for outcomes outside of the cases that inspired them. Furthermore, there will be a fundamental conflict between allowing unlimited additions of variables and maintaining assertiveness and generality in prediction. Additional information will be necessary to determine the value of each added independent variables, and such information may be available only in a narrow range of settings. Leaving such variables open, however, will make predictions of the theory unclear. Another practical consequence is that it is likely that a great multitude of versions of a single theory will be in circulation at any single time, making it unclear which one should be applied to any particular circumstances.

A way of combining the arguments for parsimony with those for generality is to note that the gap between the number of cases being explained and the number of relevant variables in a theory corresponds to the degrees of freedom in a statistical model being applied to particular empirical outcomes. Given that fewer degrees of freedom increase the extent to which accurate “predictions” can be generated for existing data via the mere manipulation of coefficients regardless of the particular independent variables used, there is typically a tendency to question any statistical model in which the number of variables is too large relative to the number of cases.

The extent to which the assumptions of a theory can be manipulated to fit existing data depends on both the narrowness of scope conditions and the complexity of the theory, and the mere empirical correspondence of a theory’s predictions with observed outcomes is not a sufficient reason to be confident that it is uncovering the underlying causal mechanisms behind those outcomes. The more general and parsimonious a theory, the more likely it is to be a “deep” explanation rather than an exercise in after the fact curve fitting. Hence, critics often argue that midrange models engage in extrapolation rather than an attempt to uncover underlying processes. Eckstein, for instance, refers to the most extreme cases of this as

\(^{25}\) Popper 1958, 42.
"mindlessly technical work" that "convert(s) data automatically into theory."  

Of course, when theories based on a general, parsimony model are shown to generate inaccurate predictions compared to theories with narrower scope conditions, it is important to incorporate the insights of the latter theories into the model in a way that does not make excessive sacrifices with regard to generality, parsimony and assertiveness. Nor should one ignore inaccuracies of assumptions any more than inaccuracies of prediction. Indeed, the former are often linked to the latter. Where an actor model’s assumptions are shown to be fundamentally wrong, this generally means that there is some flaw in the way it is modeling the causal relationship between structural circumstances and human behavior, and is likely to lead to flawed prediction.

While a concern with underlying processes behind observed empirical relationships is often seen as one of the strengths of the conventional rational choice model, rational choice theorists have often been criticized by behavioral decision theorists and other empirical researchers for ignoring evidence about individual-level processes when such evidence goes against its assumptions.

Indeed, the conventional rational choice model’s inaccurate assumptions are best seen as a compromise between a concern for modeling process and the need to generate assertive predictions, rather than an uncategorical asset to be celebrated. Contrary to Friedman’s celebrated argument mentioned at the beginning of the book, the inaccurate assumptions of purely materialistic preferences, purely logical belief formation, and purely optimizing decision-making can hardly be credited with increasing the accuracy of the model. After all, the predictive implication of these assumptions is that individuals will never do anything that can be shown from available information to be suboptimal for maximizing material welfare. This is hardly an accurate prediction.

On the other hand, making such clear assumptions is more likely to generate falsifiable predictions in the first place than leaving preferences, beliefs, and decision-making processes vaguely specified. Hence, in making evaluations of theoretical usefulness, it is important to separate the asymptotic accuracy of assumptions from accuracy in each detail, or, in other words, to look at both the assertiveness and the accuracy of assumptions as crucial to the assertiveness and accuracy of predictions. It is true that no individual matches the assumptions of the conventional rational choice model. However, as noted, where divergences from the model cannot be stated in a general, parsimonious manner, simply arguing that preferences can be nonmaterialistic, beliefs illogically determined, or decision making suboptimizing does not provide the basis for making better predictions. When divergences can be be modeled in such a fashion, there is in

---


27 Riker 1990, 176.

28 See Friedman 1953.
principle no reason while they cannot incorporated into a reformed actor model, as I have attempted to do in the preceding chapters. The main question is not whether any set of general actor assumptions can be used to accurately explain all observed behavior (they can’t, without bizarre structural assumptions) but rather whether it is possible to generate new general and parsimonious assumptions that can account for the phenomena that the conventional rational choice model cannot.

Ultimately, there is no set of criteria that we can use to judge the all-purpose relative importance of generality and parsimony vis-à-vis closeness of fit to existing data. No general, parsimonious model can avoid significant anomalies when applied to widely varying environmental contexts. On the other hand, no collection of narrow empirical observations or even midrange theories can provide the coherence and cumulability of a general model. Rather than trying to adjudicate, it is more important to see how general and midrange theorizing can be seen as complementary rather than competitive. Although this may be somewhat of a truism, it is important to emphasize the key role that each type of theorizing plays in the advancement of the other. Aggregation of midrange theories and prediction in cases in which it is difficult to make inductive measurements of decision-making processes requires the construction of general models. For those general models to be accurate, however, they must be informed by the regularities in behavior turned up by empirical observation and midrange theorizing. Theory aggregation at any rate is a process that inevitably takes place from both the top down and from the bottom up, and the main concern for social scientists is to make sure that the two paths are converging rather than diverging.

Summing Up
The limits and usefulness of models in the social sciences are illustrated by the following quotes:

For it is an illusion to think that we can reach the complexity of the real as such, and that is why the constructs we call models are indispensable tools of knowledge. It is also why reality cannot be contained in them.  

and

social scientists create robust structures in the knowledge that they may be standing on quicksand. And still they go on, because this is the only way to create some order in disorder and to discern pattern, without denying the ultimate uncertainty that makes the study of human affairs so exasperating yet so tantalizing.

The quotes reveal the tentative and often haphazard nature of theoretical innovation, but also the need to press on regardless. It with this motivation that this book was first conceived and subsequently carried out. I hope that similar motivations will cause those who

29 Boudon 1986, 227.
find shortcomings in the work to put forward their own efforts against the current state of theoretical gridlock rather than accepting it as inevitable.
References


References

References


References


References


References


References


References


References 291


References


References


References


References


References


Simmel, Georg. 1908. *Sociology: Studies of the Forms of Societalization*.


References


References


References


