COPING WITH COMPLEXITY:

ESSAYS ON EVOLUTION AND INSTITUTIONS

by

Alexander Schaefer

A Dissertation

Submitted to the Faculty of the

DEPARTMENT OF PHILOSOPHY

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2021
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: Alexander Schaefer
titled: Coping with Complexity: Essays on Evolution and Institutions
and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

David Schmidtz
Date: Jun 9, 2021

Justin Bruner
Date: Jun 9, 2021

Thomas Christiano
Date: Jun 9, 2021

Vlad Tarko
Date: Jun 11, 2021

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

David Schmidtz
Philosophy
Date: Jun 9, 2021
Acknowledgments

I am immensely grateful to my committee members — Justin Bruner, Vlad Tarko, and Thomas Christiano — all of whom devoted their valuable time to give me invaluable feedback on these essays. I’ve also had the fortune of working with two great mentors who were both advisors to this project at different times: Jerry Gaus and Dave Schmidtz. Finally, my friends, especially Mario Iván Juárez García and Matt Jeffers, provided the kind of harsh but honest feedback typically obtainable only from true friends or unkind strangers. Obviously, I also owe great debts to many people who never read or provided feedback on these chapters. Most importantly, my parents.

Finally, I’d like to thank the ETD administrators for imposing a series of arbitrary formatting requirements on my dissertation. Coping with such tedium truly builds character.
# Table of Contents

## Abstract

1. **Introduction** .................................................. 7

---

## Hayek on Complexity and Evolution

2. **Coping with Complexity: A Theory of Hayekian Interventionism** .......... 9
   2.1 Introduction .................................................. 10
   2.2 Hayek contra Intervention: 3 Central Arguments ....................... 11
   2.3 The Puzzle of Hayekian Intervention .................................. 15
   2.4 Complexity: The Unifying Concern ..................................... 18
   2.5 Hayekian Intervention as a Complexity Approach ....................... 21
       Key Themes in Contemporary Complexity Policy ...................... 22
       Hayekian Interventionism ........................................... 25
   2.6 Addressing the Puzzle .............................................. 29
   2.7 Conclusion ..................................................... 32

---

3. **Hayek’s Twin Ideas: Reconciling Methodological Individualism and Group Selection** .......................................................... 34
   3.1 Introduction ..................................................... 34
   3.2 The Inadequacy of CAS1 and CAS2 Approaches ......................... 36
       Downward Causation: Hayek’s Evolutionary Theory ................... 36
       Upward Causation: Methodological Individualism and Spontaneous Order 38
   3.3 Reuniting the Twins: Hayek’s Multilevel Selection Theory ............. 42
   3.4 Conclusion ..................................................... 46

---

## Polycentric Governance

4. **Polycentricity and Adaptation: A Multilevel Selectionist Approach** ........ 48
   4.1 The Puzzle of Polycentric Functionality ................................ 48
   4.2 The Price Equation and Multilevel Selection ............................ 50
       The Basic Price Equation ........................................... 50
       The Cultural, Multilevel Price Equation ................................ 52
       Implications of the Price Equation ................................... 53
       Group Fitness and Group Welfare ..................................... 56
4.3 Polycentricity and Multilevel Selection ........................................... 57
  Defining Polycentricity .......................................................... 57
  Facilitating Multilevel Selection .............................................. 58
4.4 Examples: Success and Failure ..................................................... 59
  CAMPFIRE Project Zimbabwe .................................................... 60
  Ostrom’s Cases of Failure .......................................................... 60
  The Importance of Monitoring and Enforcement ......................... 61
4.5 Fortifying Polycentric Governance ............................................... 61
4.6 Conclusion ................................................................................. 63

5 The Problem of Complexity and the Emergence of Polycentric Political Order 64
  5.1 Introduction ............................................................................ 64
  5.2 Polycentric Political Orders ...................................................... 65
  5.3 The Complexity Problem ......................................................... 66
    What is Complexity? ............................................................... 66
    Social Complexity and Direct Control ................................... 67
    Social Complexity and Polycentricity ................................... 69
    A Simple Model of Governing Complexity .......................... 71
    Implications of the Model ....................................................... 72
  5.4 Empirical Analysis ................................................................. 74
    Social Development and the Seshat Data Set ......................... 75
    Emerging from Feudalism ....................................................... 76
  5.5 Conclusion ................................................................................. 83

Dynamic Justice .............................................................................. 84

6 Is Justice a Fixed Point? ................................................................. 85
  6.1 Introduction ............................................................................ 85
  6.2 Justice as a Stable Equilibrium ............................................... 86
  6.3 The Problem of Existence ....................................................... 89
  6.4 Dynamic Theorizing .............................................................. 98
  6.5 Conclusion ................................................................................. 102

Appendix .......................................................................................... 104

A Appendix ...................................................................................... 104
  A.1 Deriving Price’s Equation ...................................................... 104
  A.2 Deriving the Multilevel Price Equation ................................ 107
  A.3 Henrich’s Defense of Cultural Group Selection ...................... 114

References ....................................................................................... 117
List of Figures

3.1 Multilevel, bidirectional causation ........................................ 45
5.1 Relative effectiveness of direct control and polycentric paradigms .... 72
6.1 The well-ordered society is a stable fixed point, despite constant institutional change. ......................................................... 88
6.2 Two fixed points exist, one at (0, 0) and one at (.5, .5). .................... 91
6.3 Two fixed points exist, one at (0, 0) and one at (1, 1). ...................... 92
6.4 By violating (C1), we have eliminated the fixed points at (0, 0) and at (1, 1). 93
6.5 By violating (C3), a discontinuous transition function can exist that does not exhibit any fixed points. .............................. 95
6.6 A multivariable system can exhibit cycles, one of which is indicated by the dark circle. By violating (C4), we can depict this system as lacking fixed points. 97
A.1 The Partition Theorem Illustrated by a Truly Great Artist ................ 105
A.2 Population Dispersion of z-trait ................................................. 110
A.3 Population Dispersion of Fitness .................................................. 111
A.4 z-Dispersion Between Groups .................................................... 111
A.5 z-Dispersion Within Groups ....................................................... 112
A.6 Relating partial regression coefficients and fitness ......................... 115
Abstract

The challenges this work addresses stems from the complexity of contemporary society, which imposes hard limits on our ability to predict, and thus control, social outcomes. After analyzing society as a complex adaptive system (part I), I consider the role of the state within such a system, arguing for a polycentric, democratic governance structure that permits order to arise through rivalrous, evolutionary processes (part II). Finally, on this social scientific foundation, I critically evaluate current discussions of justice in the philosophical literature (part III). I reject the popular idea that a just society is best conceived of as a stable fixed point. Rather than a static equilibrium, I argue, justice in a complex society must be understood as an evolutionary process.
Despite their disparate subjects, the following essays share a number of common themes. Chief among these are complexity, evolution, and institutions. The first two essays examine Hayek’s social theory, an examination that brings to light two basic points about complex societies. First, they are difficult to predict and control. Second, they adapt to internal and external changes.

These features lay the groundwork for section II, which examines the proper form of governance structures for a complex, adaptive society. The first essay of part II applies multilevel selection theory to the problem of governing complexity. It concludes that polycentric political organization, supplemented by a few additional “design principles,” facilitates a socially beneficial process of competition and evolution. The second essay of part II uncovers a related benefit of polycentric governance. Due to its decentralized and competitive nature and due to the vast amount of relevant and constantly changing information generated in a complex society, polycentric governance institutions utilize information more effectively than centralized modes of governance. There is a substantial welfare benefit to utilizing this information by implementing reforms that seek to address the concerns and satisfy the preferences of millions, or perhaps billions, of citizens. Moreover, centralized governance becomes increasingly difficult as increasing numbers of increasingly interdependent variables become relevant to any given problem. Polycentricity is an adaptation of the state in response to the problem of social complexity.

Finally, part III begins to examine some of the normative, philosophical consequences of these social scientific investigations. If society is in a constant state of flux, if it is evolving in response to fluctuating variables, then the traditional task of political philosophy may stand in need of amendment. Philosophers from Plato to Rawls have attempted to characterize a conception of justice, a political sumnum bonum, that transcends the institutional variations of time and place. If justice is, at least in part, a project of reconciliation, and if the values held by citizens continue to evolve, then there may not be a stable conception of justice that transcends societal dynamism. We may, instead, need to content ourselves with identifying certain general desiderata that better enable society to coordinate on a shared conception of justice, however ephemeral this conception might prove to be.
HAYEK ON COMPLEXITY AND EVOLUTION
Before he had finished speaking to his paper, [Thatcher] reached into her briefcase and took out a book. It was Friedrich von Hayek’s The Constitution of Liberty. . . . ‘This,’ she said sternly, ‘is what we believe,’ and banged Hayek down on the table."

2.1 Introduction

Margaret Thatcher was interrupting a member of the Conservative Research Department who was advocating a policy approach that blended free-market conservatism with progressive economic controls. This vignette, whether true or not, underscores a common misperception about Hayek’s stance on economic intervention. Contrary to widespread opinion, F.A. Hayek supported active governmental involvement in the economy. In fact, according to Hayek, the principle of laissez-faire has likely “done more harm to the liberal cause” than any other idea (Hayek 2007, 71; see also Hayek 1980c, 110). When it comes to concrete policy proposals, Hayek’s writings suggest ambivalence towards government programs, and he sometimes supports even programs that are vehemently opposed by other advocates of limited government.\(^1\). In the more abstract areas of governmental activity, the creation of laws for example, Hayek’s view is highly nuanced, recommending the use of both “spontaneous ordering forces,” as well as direct legislative intervention (Hayek 1983, 89).

Nevertheless, Hayek deserves his credentials as an advocate of free-markets and limited government. His enthusiasm for the market order, or catallaxy, persists as a common theme throughout his massive oeuvre. What makes the catallaxy a “marvel,” according to Hayek, is its ability to solve the fundamental problem of economics: utilizing knowledge that is widely dispersed, often inarticulate, and therefore inaccessible to any individual or organization (Hayek 1980c, 77-91). Moreover, Hayek also recognizes the existence of other such “spontaneous orders,” for example, law and morality. Intervention disrupts spontaneous order — at least in some narrow sense of intervention — undermining the use of decentralized mechanisms of coordination, such as the catallaxy.

On the one hand, therefore, Hayek views intervention into decentralized orders, such as the catallaxy, as inimical to their healthy functioning (Hayek 1980c, sec. VI). On the other hand, he envisions an important role for government in managing a successful economy, and perhaps, by extension, a successful society.\(^2\) Reconciling these two aspects of Hayek’s thought poses an interesting challenge, one that runs into various contemporary debates regarding the interpretation of Hayek’s position.

---


\(^2\) Given Hayek’s rejection of an overarching goal for society, exactly what counts as “success” is a controversial question. Eric Mack suggests that Hayek provides a “telic” justification of rules and institutions in which they aim at an abstract order of actions, as opposed to a concrete outcome. Success would then mean the successful facilitation of this spontaneous order. See (Mack 2006).
Rejecting the claim that Hayek was careless and inconsistent in his dual advocacy of spontaneous order and active government, this paper aims to provide a reading that reconciles these two aspects of Hayek’s thought. While several passages in Hayek pose a challenge to interpreting his stance on intervention, recognizing Hayek’s theory of complex systems as a foundational element of his social and political thought harmonizes a great deal of apparently conflicting claims. It thereby clarifies a Hayekian theory of intervention. Hayekian interventions are those that take full account of the epistemic challenges posed by complex systems, like society. By contrast, the interventions that Hayek most vehemently opposes exhibit a naive hubris with regard to the management of such systems by treating them as if they were similar to simple or “unorganized” systems — i.e. closed systems which contain either few parts or low levels of interdependence between the parts. The aim of this paper is not to defend Hayek’s position in all of its particulars or to point out its shortcomings. Rather, I aim to shed light on one fundamental aspect of his approach, his theory of complexity. This aspect resolves certain tensions in Hayek’s position and thus reveals a coherence in Hayek’s view that might otherwise go unnoticed.

In the next section, I will identify three of Hayek’s arguments against intervention, which I take to be his central objections to intervening in decentralized orders. Section III then presents some textual evidence that Hayek supported interventions of various kinds and clarifies the puzzle of reconciling spontaneous order with active intervention. Beginning in section IV, I develop a complexity reading of Hayek’s stance on intervention, first arguing (sec. IV) that Hayek’s arguments against active intervention are grounded in his theory of complex systems, then showing (sec. V) that Hayek’s support for interventions of certain types also has a basis in his theory of complexity. In particular, section V proceeds by presenting close parallels between Hayek’s remarks on policy reform and the policy approaches advocated by contemporary complexity theorists. Before concluding, section VI applies these complexity considerations, revealing a coherent solution to the puzzle of section III, that is, the puzzle of Hayekian intervention.

2.2 Hayek contra Intervention: 3 Central Arguments

“Intervention” is used broadly in this paper. It includes any conscious effort to affect emergent, system-level properties. Governmental central planning of production and consumption is a form of intervention, since it supersedes the prices and decisions that emerge from the market process in order to achieve a particular result on a systemic scale. Central planning within firms is not intervention, at least, so long as the goal is to maximize firm profits, rather than to affect some property of the social order as a whole. The distinction is not between public and private: judge-made law is not intervention when judges restrict themselves to articulating the rules of a pre-existing order, and firms might engage in intervention when they lobby for special laws, such as tariffs or subsidies, that will affect the social order as a whole, or even when they fund private campaigns that aim to alter the social order in some significant way. The aim of this paper is not to defend Hayek’s position in all of its particulars or to point out its shortcomings. Rather, I aim to shed light on one fundamental aspect of his approach, his theory of complexity. This aspect resolves certain tensions in Hayek’s position and thus reveals a coherence in Hayek’s view that might otherwise go unnoticed.

In the next section, I will identify three of Hayek’s arguments against intervention, which I take to be his central objections to intervening in decentralized orders. Section III then presents some textual evidence that Hayek supported interventions of various kinds and clarifies the puzzle of reconciling spontaneous order with active intervention. Beginning in section IV, I develop a complexity reading of Hayek’s stance on intervention, first arguing (sec. IV) that Hayek’s arguments against active intervention are grounded in his theory of complex systems, then showing (sec. V) that Hayek’s support for interventions of certain types also has a basis in his theory of complexity. In particular, section V proceeds by presenting close parallels between Hayek’s remarks on policy reform and the policy approaches advocated by contemporary complexity theorists. Before concluding, section VI applies these complexity considerations, revealing a coherent solution to the puzzle of section III, that is, the puzzle of Hayekian intervention.

“Intervention” is used broadly in this paper. It includes any conscious effort to affect emergent, system-level properties. Governmental central planning of production and consumption is a form of intervention, since it supersedes the prices and decisions that emerge from the market process in order to achieve a particular result on a systemic scale. Central planning within firms is not intervention, at least, so long as the goal is to maximize firm profits, rather than to affect some property of the social order as a whole. The distinction is not between public and private: judge-made law is not intervention when judges restrict themselves to articulating the rules of a pre-existing order, and firms might engage in intervention when they lobby for special laws, such as tariffs or subsidies, that will affect the social order as a whole, or even when they fund private campaigns that aim to alter the social order in some significant way. The aim of this paper is not to defend Hayek’s position in all of its particulars or to point out its shortcomings. Rather, I aim to shed light on one fundamental aspect of his approach, his theory of complexity. This aspect resolves certain tensions in Hayek’s position and thus reveals a coherence in Hayek’s view that might otherwise go unnoticed. 
Hayek levels many objections against interfering in the voluntary activities of individuals and organizations within society in order to achieve system-level goals. These objections are often related by key themes in Hayek’s work, such as the dispersal of information or the requirements of spontaneous order. The three arguments presented here provide one way to sort out Hayek’s most salient objections to intervention. While this categorization is clarifying, its artificiality becomes apparent in the tight connections between these arguments. In fact, as section IV will argue, all three find support in Hayek’s theory of complexity, which, on my reading, is the logical foundation of Hayek’s views on intervention.

(i) The Argument from Diversity and Authority: general rules allow cooperation among a diverse populace that could not agree on reasons for intervening that violate these rules.

This argument, which constitutes the core of the Road to Serfdom, turns on a contrast between systems based on centralized planning and systems based on unplanned orders, which emerge spontaneously from general rules, both formal and informal. In The Road to Serfdom, Hayek develops an impossibility proof, seeking to demonstrate an incompatibility between the values affirmed by socialist planners and the methods they propose to achieve those values. Without denying that their goals are laudable, Hayek claims that the methods they propose to attain these goals are unacceptable, not just to the opponents of socialism, but to socialists themselves. To be more specific, Hayek’s socialist affirms the freedom and equality of each individual, on the basis of which he or she advocates a planned society that could ensure a just distribution of economic surplus. However, central planning requires, by definition, centralized direction of production and consumption. It thus removes such decisions from the sphere of personal choice and relocates them to an official body that will determine which purposes are worthy to receive support and funding and which are not. Consequently, such a system manages people and their productive energies as resources for the achievement of purposes which they may or may not recognize. When economic decisions are to be made by a central body, then it is this body, rather than individuals, that must determine the relative importance of diverse and incompatible ends. This, Hayek argues, entails a rejection of freedom and an affirmation of unequal partiality (Hayek 2007, 128, 130).

Notice the role of diversity in this argument: individuals have diverse and incompatible preferences, yet in order to bring about its ends, a planning authority must “reduce the diversity of human capacities and inclinations to a few categories . . . and . . . disregard minor personal differences” (ibid., 130). Modern, large-scale societies are full of individuals with diverse dispositions and competing values. Among individuals in such a society, “there will exist no agreement on the relative importance of their respective ends,” and consequently there is no agreed-upon criteria by which to evaluate social outcomes (Hayek 1997, 201-2). Tellingly, Hayek suggests that when society faces an existential threat, such as a major war with an aggressive and powerful opponent, the shared and overriding interest of survival softens the moral predicament of central planning (e.g. ibid., 122, 168). Our individual differences are superseded by a single goal. And when we are told to produce more guns and consume less butter in order to serve this goal, we will not object that we are being used for the purposes of others.

4: Notice that Hayek usually speaks in terms of values or inclinations, rather than preferences. The term “preferences” has the advantage of avoiding the psychological or ethical baggage that Hayek’s terms lug with them.

5: For an excellent discussion on this issue, see (Mack 2006, 261, 271, 274). See also (Gaus 2018, 5).
Encapsulating this argument, Hayek asserts that (1) socialist planners value freedom and equality, but that (2) carrying out socialist plans (which dictate production and consumption decisions) is incompatible with respecting these values. Therefore, (3) society must choose between either entrusting some authority to determine the use of social resources or upholding the kind of freedom and equality that is only possible when individuals make their own decisions about how to use the resources that they produce or acquire.\(^6\)

(ii) The Knowledge Argument: knowledge is dispersed and tacit, yet necessary for rational and efficient coordination, and only a spontaneous order is able to leverage such knowledge.

Hayek is pointing to the sheer difficulty of acquiring the knowledge necessary to produce a rational economic plan for society — i.e. one that satisfies feasibility and efficiency. In particular, he draws our attention to the nature of the information that would be required to determine the relative trade-off ratios (viz. prices) between various goods and services. In essence, Hayek’s argument is that the data required to determine rational trade-off ratios — data about the beliefs and preferences of individuals, as well as of special skills of production — is always dispersed throughout society, often unknown to any single individual, and (most importantly) often lies tacit and inarticulable in its possessor. Moreover, the subjective data are constantly in flux, since interactions between individuals within the market process shape their beliefs and preferences (Hayek 1961). Consequently, the data can never be gathered and entered into the equations that would allow us to calculate equilibrium prices and quantities.\(^7\)

The error of the prospective planner lies partly in the hubristic effort to supersede the existing rules, which constitute tacit knowledge (Hayek 2011, 77). It lies also in the implied subordination of dispersed centers of decision-making to the decisions of the planning authority. Such subordination reduces the amount of knowledge available in making plans to the limits of a single consciousness (or to the limits of a planning board). Due to the dispersed and tacit nature of the data, these limitations are severe; economic planning conducted in this manner will be under-informed and, consequently, irrational (Hayek 1980c).

(iii) The Argument from Predictive Difficulty: Society is unpredictable and thus uncontrollable.

Hayek’s third argument against central planning rests on a distinction, first expressed in his 1955 “Degrees of Explanation,” that marks a major advance in his thought: that between simple phenomena and complex phenomena.\(^11\) Simple phenomena are systems composed of relatively few parts with relatively weak connection between these parts. Phenomena of “organized complexity,” on the other hand, are systems containing a large number of connected and interdependent elements.\(^8\) The character of such phenomena depends not only “on the properties of the individual elements of which they are composed, and the relative frequency with which they occur, but also on the manner in which the individual elements are connected with each other” ([1975] 2014, 365). Such systems exhibit strong feedback relations between their elements, resulting in special properties, such as sensitivity to initial conditions, network structures, and quantities.

6: To complete the argument, Hayek spends much of the book presenting reasons to prefer this kind of freedom and equality to the ends that socialist planning might successfully achieve. One such reason is the difficulty that a planner would have in actually achieving these ends, meaning that our freedom and equality will have been sacrificed for comparatively little. This aspect of the argument is discussed under the third category of arguments, the “prediction-control argument.”

7: As Hayek ([1975] 2014) explains: “It is true that their systems of equations describing the pattern of a market equilibrium are so framed that if we were able to fill in all the blanks of the abstract formulae, i.e., if we knew all the parameters of these equations, we could calculate the prices and quantities of all commodities and services sold. But as Vilfredo Pareto, one of the founders of this theory, clearly stated, its purpose cannot be ‘to arrive at a numerical calculation of prices,’ because, as he said, it would be ‘absurd’ to assume that we could ascertain all the data,” (366).

8: Hayek identifies a third category of phenomena, those of “unorganized complexity,” which, though comprised of many entities, can be analyzed purely in terms of (1) the properties of the entities the phenomena comprise, and (2) “the relative frequency” of these entities. Such phenomena can be studied using standard statistical methods (Hayek [1975] 2014, 365).
and path dependencies. These properties undermine the possibility of precise predictions. For example, the relative benefits of a given phenotypic expression of a gene depend upon a massive number of other factors: the traits exhibited by other members of the same species, the traits exhibited by species inhabiting the same environment, the structure of interactions between various organisms, even random events that affect organisms’ fitness. Moreover, one organism’s response to its environment constitutes part of the environment to which connected organisms must respond. When the relative fitness of traits exhibit such interdependencies, it is impossible to locate an optimal set of traits for an organism to exhibit, for by the time such an optimal set of traits is attained, the environment will have changed, likely rendering such a set suboptimal. In the terms of modern complexity theory, the fitness landscape is “dancing” (Page 2011, 93-4). It is thus impossible to predict the precise set of traits toward which a species will gravitate, since this depends upon the massive number of reactive adjustments made along the way by other organisms inhabiting the same environment.

Although scientists recognize that it would be hopeless to apply the theory of evolution with the hopes of predicting the future genetic make up of a given species, Hayek believes that economists fall into a similar error when they attempt to use economics to make precise predictions about society. As a phenomenon of organized complexity, a social system gives rise to events that “depend on so many concrete circumstances that we shall never in fact be in a position to ascertain them all” (Hayek [1964] 2014, 269). Economists, in Hayek’s view, ignore the complexities of social systems and, consequently, treat them as simple or unorganized systems, importing inappropriate methods from the physical sciences that lead them into serious errors. This error leads social scientists to believe that they can predict and control social systems — yet, if societies are complex as Hayek argues, the supposed knowledge that prediction and control requires is mere pretense.

To summarize the argument from predictive difficulty: (1) Only simple or unorganized systems are amenable to successful (precise) prediction, (2) Society is not a simple or unorganized system; it exhibits organized complexity. (3) Society is thus not amenable to successful (precise) prediction. (4) Successful (precise) control implies (precise) prediction. Therefore, (5) society cannot be successfully (precisely) controlled.

At this point, two caveats are in order.

First, Hayek does not claim that all prediction is impossible in systems of organized complexity, and therefore does not claim that all control is impossible. Rather, we are restricted to what he calls “pattern predictions,” or equivalently, explanations in terms of general principles. Therefore, only interventions that depend on the accuracy of precise or long-term predictions are ruled out by his argument. This will become important when we turn to developing a theory of Hayekian intervention.

Second, unpredictability does not imply undesirability. Suppose, for example, that intervening into the normal functioning of social rules or the market process causes a brilliant scientist to direct her efforts toward developing a new technology that produces clean, renewable energy at stunningly low cost. Without such an intervention, imagine, she would have developed a new superweapon, instead, or perhaps a new shoe
Coping with Complexity

A Theory of Hayekian Interventionism

The spontaneous order arises from each element balancing all the various factors operating on it and by adjusting all its various actions to each other, a balance which will be destroyed if some of the actions are determined by another agency on the basis of different knowledge and in the service of different ends” (Hayek 1983, 51).

One response is simply to say that if there is little to no prospect of achieving one’s aim, then there is no point in incurring the costs of intervention. However, there is another response to this caveat, one which rests on Hayek’s notion of spontaneous order. Like the organisms in an ecosystem, human beings and their various plans exhibit strong interdependencies. Mutual expectations and joint compatibility of diverse actions constitute a delicate equilibrium resting on a set of formal and informal rules of conduct. Furthermore, the effectiveness of any given rule depends upon the other rules in place, just as the adaptiveness of any given species trait depends upon the traits of other organisms within the ecosystem. Altering social rules is thus disruptive in multiple ways. First, it disrupts the balance between expectations and actions in the same way that removing or adding a species to an ecosystem will disrupt the balance of the ecosystem. Second, it may alter or abrogate rules that contribute to the effectiveness of the other rules in place. Although it is certainly possible that intervening will leave an ecosystem intact, or even healthier, the fact that many species depend upon the continued, predictable activities of other organisms in their environment implies that altering ecosystems is generally harmful. Similarly, altering or abrogating rules generates confusion and disequilibrium, justifying Hayek’s presumption against intervention in the absence of reliable predictions.

We have now completed our survey of Hayek’s most salient arguments against intervention. It remains to catalogue the various interventions he advocates (section III) and to consider how we might reconcile Hayek’s aversion to intervention with the various interventionist measures he advocates.

2.3 The Puzzle of Hayekian Intervention

As we have seen, Hayek presents a battery of challenges to the prospective central planner who wishes to control society by interfering in the lives and activities of individuals. Doing so clashes with our basic values of equality and liberty among a diverse populace, presumes the possession of unattainable information, and naively posits the ability to make precise predictions about the operation of complex systems. Arguments such as these support a typical reading of Hayek in which, as one commentator puts it, Hayek “directs his objections not only against attempts to ‘organise’ in a total or ‘utopian’ way but also against more modest ‘interferences’ with the order, which he alleges, always disrupt it,” (Vernon 1979, 64). If this is correct, then as James Buchanan puts it, Hayek’s view implies that “any ‘constructively rational’ interferences with the ‘natural’ processes of history are . . . to be studiously avoided. The message seems clear: relax before the slow sweep of history,” (Buchanan 2001, 312; see also, Buchanan 2000, 211). Yet, as I have already asserted, and as Buchanan recognized, Hayek’s view is not so simple. In his most anti-interventionist work, The Road to Serfdom, Hayek complicates his image as a libertarian

polish. Of course, we could imagine exactly the opposite scenario, in which intervention sacrifices the energy source for the shoe polish, but if interventions can produce incredibly good or incredibly bad results, then why the presumption against such interventions?

Of course, we could imagine exactly the opposite scenario, in which intervention sacrifices the energy source for the shoe polish, but if interventions can produce incredibly good or incredibly bad results, then why the presumption against such interventions?

One response is simply to say that if there is little to no prospect of achieving one’s aim, then there is no point in incurring the costs of intervention. However, there is another response to this caveat, one which rests on Hayek’s notion of spontaneous order. Like the organisms in an ecosystem, human beings and their various plans exhibit strong interdependencies. Mutual expectations and joint compatibility of diverse actions constitute a delicate equilibrium resting on a set of formal and informal rules of conduct. Furthermore, the effectiveness of any given rule depends upon the other rules in place, just as the adaptiveness of any given species trait depends upon the traits of other organisms within the ecosystem. Altering social rules is thus disruptive in multiple ways. First, it disrupts the balance between expectations and actions in the same way that removing or adding a species to an ecosystem will disrupt the balance of the ecosystem. Second, it may alter or abrogate rules that contribute to the effectiveness of the other rules in place. Although it is certainly possible that intervening will leave an ecosystem intact, or even healthier, the fact that many species depend upon the continued, predictable activities of other organisms in their environment implies that altering ecosystems is generally harmful. Similarly, altering or abrogating rules generates confusion and disequilibrium, justifying Hayek’s presumption against intervention in the absence of reliable predictions.

We have now completed our survey of Hayek’s most salient arguments against intervention. It remains to catalogue the various interventions he advocates (section III) and to consider how we might reconcile Hayek’s aversion to intervention with the various interventionist measures he advocates.

2.3 The Puzzle of Hayekian Intervention

As we have seen, Hayek presents a battery of challenges to the prospective central planner who wishes to control society by interfering in the lives and activities of individuals. Doing so clashes with our basic values of equality and liberty among a diverse populace, presumes the possession of unattainable information, and naively posits the ability to make precise predictions about the operation of complex systems. Arguments such as these support a typical reading of Hayek in which, as one commentator puts it, Hayek “directs his objections not only against attempts to ‘organise’ in a total or ‘utopian’ way but also against more modest ‘interferences’ with the order, which he alleges, always disrupt it,” (Vernon 1979, 64). If this is correct, then as James Buchanan puts it, Hayek’s view implies that “any ‘constructively rational’ interferences with the ‘natural’ processes of history are . . . to be studiously avoided. The message seems clear: relax before the slow sweep of history,” (Buchanan 2001, 312; see also, Buchanan 2000, 211). Yet, as I have already asserted, and as Buchanan recognized, Hayek’s view is not so simple. In his most anti-interventionist work, The Road to Serfdom, Hayek complicates his image as a libertarian
crusader in various passages. For instance, he rejects the terminology and the policy of “laissez-faire,” he accepts regulation of various industries (citing concerns with safety and sanitation), and he even goes so far as to claim that free-market competition is compatible “with an extensive system of social services — so long as the organization of these services is not designed in such a way as to make competition ineffective over wide fields,” (Hayek 2007, 86-7). Some might read Hayek as pandering to centrist readers, since The Road to Serfdom was, after all, an attempt to influence public opinion. Hayek did, after all, dedicate it to “socialists of all parties.” However, Hayek goes even further in his 1947 address to the Mont Pelerin Society, a venue that attracted an anti-socialist audience. In this address, Hayek entertains or endorses an surprising array of government programs: sanitation and health services, monetary management (ideally via automatic mechanisms), welfare provision for the unemployed, city planning, intellectual property, antitrust laws and regulations on the size of corporations, temporary restrictions or regulations on international trade, and an inheritance tax to support social mobility (Hayek 1980e, 109-18). Some of these policies may be regarded as unfortunate concessions that Hayek makes in order to pursue the greater task of advancing an agenda of liberty. This address might be read as a pragmatic policy approach, rather than an outline of an ideally free society. Yet, even in his abstract and thoroughly unpragmatic work on political order, work that involves highly idealized accounts of the formation of laws and the operation of society, Hayek provides wide scope for intervention. Chapter 2 of Law, Legislation, and Liberty contains what may be the most challenging passage in all of Hayek’s work for the anti-interventionist reading. Though long, it is worth quoting in its entirety:

... The fact that law that has evolved [spontaneously] has certain desirable properties does not prove that it will always be good law or even that some of its rules may not be very bad. It therefore does not mean that we can altogether dispense with legislation.

There are several reasons for this. One is that the process of judicial development of law is of necessity gradual and may prove too slow to bring about the desirable rapid adaptation of the law to wholly new circumstances. Perhaps the most important, however, is that it is not only difficult but also undesirable for judicial decisions to reverse a development, which has already taken place and is then seen to have undesirable consequences or to be downright wrong. The judge is not performing his function if he disappoints reasonable expectations created by earlier decisions. Although the judge can develop the law by deciding issues which are genuinely doubtful, he cannot really alter it, or can do so at most only very gradually where a rule has become firmly established; although he may clearly recognize that another rule would be better, or more just, it would evidently be unjust to apply it to transactions which had taken place when a different rule was regarded as valid. In such situations it is desirable that the new rule should become known before it is enforced; and this can be effected only by promulgating a new rule which
is to be applied only in the future. Where a real change in the law is required, the new law can properly fulfill the proper function of all law, namely that of guiding expectations, only if it becomes known before it is applied.

The necessity of such radical changes of particular rules may be due to various causes. It may be due simply to the recognition that some past development was based on error or that it produced consequences later recognized as unjust. But the most frequent cause is probably that the development of the law has lain in the hands of members of a particular class whose traditional views made them regard as just what could not meet the more general requirements of justice... But such occasions when it is recognized that some heretofore accepted rules are unjust in the light of more general principles of justice may well require the revision not only of single rules but of whole sections of the established system of case law. This is more than can be accomplished by decisions of particular cases in the light of existing procedures. (88-9)

At first glance, this passage is astonishing. How can the operation of spontaneous order be consistent with intentional and direct alteration of whole sections of the established system of case law? And if these are inconsistent, then how will such intervention not raise the issues that Hayek identifies in his arguments against intervention? Hayek appears to reject the gradual and evolutionary development of law — because it is too slow or is recognized as unjust — in favor of large-scale legislative interventions that exhibit all the serious failings of interventionist policies, some of which Hayek points out in the very same work.

In order to grasp the contours of this puzzle and the difficulties involved in developing an adequate interpretation of Hayek’s position, it may help to encapsulate and review the key claims involved, resulting in two propositions which appear inconsistent at first face:

1. Interventions cause great harm: Interventions trammel on the freedom of citizens and deny their equal status, they prevent the use and coordination of vast stores of dispersed and tacit information, and, in the absence of reliable predictions and feasible control, they generate unpredictable consequences.

2. Interventions of various sorts are compatible with or necessary for a healthy society: Hayek entertains the desirability of various centrally-administered regulations, welfare programs, and public services. He also asserts that spontaneous, judge-made law may evolve too slowly, and it may require changes too extensive — due to mistakes or systemic injustice — to rely upon the quasi-evolutionary process by which law may arise spontaneously.

Between claims (1) and (2) there exists a clear tension, but is it an outright contradiction? This pair of propositions fixes the challenge for a theory of Hayekian intervention: How do we avoid the pitfalls of intervention that Hayek established in arguments (i)-(iii) while also permitting the possibility of centrally administered policies and of direct legislative overhaul of the legal framework? How do we separate out those interventions that undermine society from those that support it?
2.4 Complexity: The Unifying Concern

To make sense of Hayek’s stance on intervention, that is, to understand why he views certain interventions as salutary while others are disastrous, we must look closely at his reasons for opposing intervention. Although I have already laid out Hayek’s most salient arguments against intervention (sec. II), there is an underlying theory that unifies these arguments and points toward a principled account of Hayekian intervention. I am referring to the theory of complex systems.

Hayek made pioneering contributions to the study of complex systems starting in 1955 with the publication of “Degrees of Explanation,” and continued to write on the topic throughout the following decades, most notably in “The Theory of Complex Phenomena,” and in “The Pretense of Knowledge.” Hayek’s characterization of complexity — as several commentators have noted (e.g. Vaughn 1999; Lewis 2017; Lewis and Lewin 2015) — mirrors contemporary characterizations of complexity. According to Scott Page, for example, complex systems are characterized by connected, interdependent, adaptive (or rule-following), and diverse entities (Page 2011, 25, 38). In virtue of their interdependencies, such systems exhibit network structures, feedback mechanisms, emergent properties, and path-dependence. In his various writings on complexity, Hayek mentions every single one of these features. Although the theory of complex systems is currently a live research program and continues to undergo important developments, invoking this theory is therefore not anachronistic.

In what follows, I will argue that complexity theory constitutes a unifying thread between Hayek’s three arguments against intervention. The point of doing so is to justify the approach of understanding Hayek’s theory of intervention in terms of his theory of complexity. Accordingly, in the following section, I apply the complexity framework to develop a theory of Hayekian intervention. Our new focus on the problem of complexity will allow us to better explain both Hayek’s general opposition to intervention and his advocacy of interventions of various types. The ultimate aim is to reconcile the apparently conflicting views that Hayek expresses toward interventions.

Hayek’s first argument against intervention turns on the existence of diversity in society — in particular, diversity of purposes, preferences, beliefs, and other subjective data. If we all share a common goal, as Hayek seems to think we do when facing an existential threat, then the diversity in our respective ends diminishes significantly. Consequently, each individual becomes willing to accept his or her ascribed role in an organized plan of action. In such situations, the leader overrides peoples’ individually chosen actions without forcing large numbers of individuals to sacrifice their ends for the ends of others, because in such rare and dire situations individuals actually share a “common end,” namely, survival. A necessary condition for Hayek’s conclusion in The Road to Serfdom, therefore, is a certain level of diversity and disagreement among individuals.

Where does this diversity come from and why is it so endemic to modern society? Contemporary complexity theorists, such as Brian Arthur, have
argued that complexity actually generates diversity (Arthur 1994, 66-9). The phenomenon of complexity occurs when several interacting elements, tightly bound by what Hayek and others have called “feedback relations,” interact and adapt to one another. As this interaction unfolds, new opportunities or “niches” arise, allowing for the entry of new and different entities—different strategies, for example. These new entrants, by interacting with already established entities, generate yet further niches, and so on, all coalescing into an upward spiral of diversity. Hayek outlines the proliferation of novel products, technologies, as well as consumer preferences, and beliefs, in his essays on the market process. For Hayek, the key to understanding this process is to grasp that “[t]he problem becomes one of how the “data” of the different individuals on which they base their plans are adjusted to the objective facts of their environment (which includes the actions of other people)” (1980d, 93). Crucially, the subjective data — preferences and beliefs — are in a state of constant flux due to “the acquisition of new knowledge by the different individuals or of changes in their data brought about by the contacts between them” (ibid., 93-4). Thus, for Hayek, as for Arthur, the interdependencies between individuals contribute to the emergence of novel beliefs, preferences, and strategies (including production methods) that drive the increasing diversity of our complex society.  

Diversity is not merely a consequence of complexity, but is also a contributing factor of complexity. Contemporary complexity theorists have explored this connection, producing various models to illuminate the effect of diversity on complexity (Page 2011, 33-41). Though Hayek never developed this relationship in detail, he does mention that complexity arises when the “the number of significantly connected variables of different kinds” is high (Hayek [1955A] 2014, 195, emphasis added). He also expresses awareness of it in his writings on market competition. In “The Meaning of Competition,” for example, Hayek makes two points that, together, link diversity and complexity: (a) competition is most important under conditions of complexity, where the outcome is unpredictable (1980d, 93-4), and (b) when various producers of a given commodity yield homogeneous products using similar methods — and, consequently, consumers have accurate and homogeneous beliefs regarding these products — “there is little need or scope for competitive activities” (1980d, 102-3). Putting (a) and (b) together suggests that we require competition in order to cope with situations where products and subjective data are diverse, because such situations are more complex than situations characterized by greater homogeneity. Complexity and diversity thus exhibit a mutually reinforcing relationship for Hayek, and for this reason complexity constitutes an important feature of his first argument against intervention.

Complexity also plays an important role in Hayek’s second objection to intervention, the knowledge argument. Recall that this argument relies on the idea that important knowledge is dispersed and tacit, i.e. that individuals know their local environments and subjective data quite well, but lack information regarding the environment and subjective data of others. What would prevent a central planner from acquiring this knowledge and using it to coordinate the plans of individuals from afar? In contrast to scientific knowledge, the knowledge informing economic decisions is often particular, rather than general, it is based on “temporary

13: As Hayek explains elsewhere: “...the tastes of man, as is also true of his opinions and beliefs and indeed much of his personality, are shaped in a great measure by his cultural environment” (1961, 347).
12: Hayek claims that society is a complex system the elements of which are complex systems. See (Hayek 1952, 19, 43ff, 185ff) and (Lewis forthcoming).
13: Page discusses a prisoner’s dilemma model that becomes complex with the addition of diverse players. The model is developed in (Nowak and May 1992).
14: Hayek does not use the word “complex” in this discussion, since it predates his use of that terminology. Nevertheless, he points the the concept that he will later call complexity by referring to the “contacts between” individuals resulting in a “process of continuous change” in those interacting individuals. He also introduces complexity into the discussion in his later essay on the market process (Hayek [1968] 2014).
opportunities” and “circumstances of the fleeting moment” (Hayek 1980c, 80), which cannot possibly be collected, disseminated to, and processed by a planner in a timely fashion. Often, such knowledge is skill-based, rather than theoretical, involving an intuitive sense of how to proceed under various contingencies and how to quickly acquire information that one does not yet possess. Consequently, the utilization of such knowledge depends upon leaving the relevant decisions to its possessor, or making them with her active participation (ibid.).

Hayek grants that achieving a complicated task, such as fighting a war or building a spaceship, poses little difficulty for an interventionist planner. The difficulty is when the task is not merely complicated, but complex. A paradigm example is that of achieving coordination or equilibrium among a vast number of individuals possessing diverse and interdependent knowledge, beliefs, and preferences. For Hayek, this task requires “the combination of fragments of knowledge existing in different minds [to] bring about results which, if brought about deliberately, would require a knowledge on the part of the directing mind which no single person can possess” (1980a, 54). Positing that centrally planned coordination faces insurmountable epistemic difficulties that arise due to the vast number of interdependent plans is tantamount to positing that the complexity of the situation — the large number of interdependent variables — renders the task of planned coordination impossibly difficult. As we have seen, in a relatively small or homogeneous group of independent individuals, the knowledge problem does not arise. According to Hayek, it is under an extensive division of labor, in which individuals' plans “require corresponding actions on the part of other individuals” (1980a, 38), that we face the difficult challenge of coordinating the beliefs and behaviors of diverse individuals. It is the complexity of modern society that gives birth to the knowledge problem.

The connection between complexity and Hayek’s third argument against intervention is the most direct of the three. The central premise is that, when dealing with society, our predictive powers face severe limitations. Society is a complex system, and predicting the behavior of complex systems require vast amounts of information:

The multiplicity of even the minimum of distinct elements required to produce (and therefore also of the minimum number of data required to explain) a complex phenomenon of a certain kind creates problems which dominate the disciplines concerned with such phenomena and gives them an appearance very different from that of those concerned with simpler phenomena. The chief difficulty in the former becomes one of in fact ascertaining all the data determining a particular manifestation of the phenomenon in question, a difficulty which is often insurmountable in practice and sometimes even an absolute one. ([1964] 2014, 263)15

When systems exhibit high levels of interdependence between a large number of elements, predicting their behavior requires ascertaining a massive amount of information about the “initial and marginal conditions” (ibid., 259). Obtaining such information is often beyond the realm of the feasible (Hayek [1975] 2014, 370). If we are limited in prediction, then we are limited in control. Pretending that we may manipulate society

15: See also (Hayek [1955A] 2014, 211): “...the more we move into the realm of the very complex, the more our knowledge is likely to be of the principle only, of the significant outline rather than of the detail. Especially where we have to deal with the extreme complexity of human affairs, the hope of ever achieving specific predictions of particulars seems vain.”
Coping with Complexity: A Theory of Hayekian Interventionism

as if it were a simple system will lead to “deplorable effects” (ibid., 368), since the attempt to control society will “impede the functioning of those spontaneous ordering forces by which, without understanding them, man is in fact so largely assisted in the pursuit of his aims...” (ibid., 371). The spontaneous order, Hayek’s alternative to central planning, is essentially a method — accidentally discovered, unwittingly employed — for coping with complexity. By undermining the spontaneous ordering forces of society, intervention by a planning authority, undermines our ability to deal effectively with the complexities arising from a diversified and highly interdependent society. Understood this way, Hayek’s view of society and intervention are fundamentally rooted in his theory of complexity. Diversity is its condition and its consequence. Coordinating a diverse and heterogeneous set of individuals, each of which possesses important, though inaccessible, knowledge is essentially a problem of complexity — that is, a problem that requires solving various connected, interdependent equations without having full access to the information such a solution would require. Furthermore, the paradigm of prediction-and-control breaks down when applied to society simply in virtue of its prediction-limiting complexity. We will now see that understanding the fundamental problem as one of complexity allows us to make sense of Hayek’s position on intervention, and allows us to reconstruct a theory of Hayekian intervention with deep affinities to the approach proposed by contemporary complexity theorists.

2.5 Hayekian Intervention as a Complexity Approach

In the previous section, I argued that Hayek’s theory of complexity is fundamental to his opposition to intervention. All three major arguments that Hayek levels against active intervention have some basis in his analysis of society as a complex system. Intriguingly, Hayek’s theory of complexity also illuminates why Hayek finds certain types and methods of intervention less objectionable. Complexity-based issues present a hurdle to effective policy. Overcoming this hurdle, though not sufficient, provides a necessary check on interventionist proposals. Contemporary complexity theorists, though their focus tends to be more narrow, share Hayek’s awareness of complexity as constraining the feasible set of effective policy interventions. To better understand why Hayek accepts some interventions as, at least, worthy of consideration while viewing others as objectionable in principle, it is valuable to examine how these complexity theorists approach policy reform and how this approach parallels Hayek’s in several ways. Although the complexity approach to policy remains underdeveloped (Colander and Kupers 2014, 6, 53), many complexity theorists have begun to explore the policy implications of viewing society as a complex adaptive system (Wilson and Kirman 2016; Colander and Kupers 2014; Axelrod and Cohen 2000). Like Hayek, these theorists emphasize the unpredictability of intervening in society and the importance of drawing on local knowledge and adjustments to bring about benefits for society as a whole. And, like Hayek, these theorists tend to emphasize the process of social evolution over the desirability of particular equilibrium states. 

16: Showing that the ideal policy is consistent with societal complexity does not show that such a policy is likely to emerge from the political process. Indeed, Hayek was a forerunner of public choice economics, often pointing out the perverse incentives that government actors face. In The Road to Serfdom, for example, Hayek argues that an ample sphere of government action will lead to interest group conflicts and will attract unsavory character types into government positions. We must therefore recognize that policy formulation and implementation are subject to the influence of interest group lobbying and to that of politicians with little to no regard for public welfare. For a discussion of Hayek and public choice, see (Leeson and Boettke 2003).

17: Especially interesting in this respect is Brian Arthur’s pioneering work on “complexity economics,” which is replete with Hayekian themes. See, for example, (Arthur 2013, 16).
The recognition that a complex society is unpredictable and dynamic has clear affinities with Hayek’s way of thinking about society and intervention. The approach to intervention and public policy that follows from this picture of society reveals further similarities between Hayek and contemporary complexity theorists. Such similarities suggest that at base it is Hayek’s theory of society as a complex system that gives rise to his particular views on intervention.

The remainder of this section will elaborate on some key themes in contemporary approaches to policy from a complexity framework. It will then show how these themes figure into Hayek’s claims about proper and improper interventions, establishing Hayek’s theory as one fundamentally concerned with the challenges and opportunities that complex systems pose for interventions of various sorts.

Key Themes in Contemporary Complexity Policy

One way of conceptualizing the nature of a complex system is to view it as a set of epistatically linked (i.e. interdependent) elements, different combinations of which facilitate different functionalities, or service characteristics, to different degrees.\(^{16}\) A non-complex problem lacks strong epistatic relationships between its variables, or else has very few variables. For example, as Fredrick Taylor famously demonstrated, designing an ideal shovel is a simple problem: starting very small, as the shovel gets larger (as we change its elements), its capacity to move materials (its service characteristic) increases, peaks, then steadily declines as size continues to grow. The problem of designing the optimal shovel generates a “Mount Fuji landscape” — one in which the parameters are unitary or non-epistatic, resulting in an easily discoverable optimum. By contrast, society involves millions of variables that interact in various ways. Limiting our attention to the economy, various goods and services form an interconnected web of inputs and outputs, and consequently the actions of each firm and of each consumer impinge upon the optimization problems of the others. Modifications to consumption patterns will affect the preferences, beliefs, and future behavior of consumers and producers. Tracing out the long-term effect of particular changes is impossible. Unlike simple problems, the landscape resulting from a complex problem is constantly shifting — as we change the values of certain variables in order to approach an optimum, this change in variables impacts other variables both by altering the values they may take and by altering the contribution to service characteristics that these variables make. In the terminology of complexity theory, the result is a “dancing landscape.”

In the work of complexity theorists who aim to address these sorts of problems, there are several recurring themes. To demonstrate the strong affinity between contemporary complexity policy and Hayek’s views on intervention, I focus on three such themes.

1. First, contemporary theorists often recommend an incrementalist and experimental approach to reform. Imposing large changes on the complex system will yield wildly unpredictable results. Consequently, making global changes is tantamount to playing roulette with the functionality of the social order. Far better is to make marginal changes to existing policy to see how they affect the functionality of the current order or to conduct small-scale experiments on rule changes before introducing them on a

---

\(^{16}\): This terminology is borrowed from Fred D’Agostino (2009).
larger scale. Smaller interventions minimize risk by generating smaller effects and by exhibiting greater reversibility.\textsuperscript{19}

In the realm of complexity policy, the incrementalist approach finds wide expression. Charles Lindblom’s famous essay, “The Science of Muddling Through” contains a classic statement of the incrementalist approach to policy, or what he calls the “branch method.” According to Lindblom, there are too many important factors with too many interdependencies to engage in an exhaustive search for the globally optimal policy. Instead, we must be content with a procedure of “successive limited comparisons” — i.e. of marginal changes, the effects of which we evaluate according to provisional standards (Lindblom 1959). More recently, evolutionary biologist and complexity theorist, David Wilson has written:

Selecting complex systems for group-beneficial outcomes is especially fraught with difficulties because interventions are likely to produce unintended consequences. According to some estimates, over half of change efforts in the business world make things worse rather than better (Schaffer and Ashkenas 2007). Given the pervasiveness of unintended consequences and cascading effects of interventions in complex social systems, there is no alternative to conducting careful experiments and scaling up practices that work... (Wilson 2016, 45).

Changes are likely to be disruptive in unpredictable ways. To protect ourselves from disaster, we must undertake only minor changes or else take steps to isolate the effects of these changes to subsections of the system.

This approach to intervention in complex systems faces important limitations. Firstly, it only applies to systems characterized by moderate complexity, rather than maximal complexity or chaos (Gaus 2018). In a maximally complex system, each element has a strong connections to every other element. Consequently, even a small change to just one of the elements will generate dramatic changes at the system level. The result is radical uncertainty as to the effect of even the most minor reforms. Furthermore, since the small change will have altered the state of all elements in the system, rolling back the reform is not feasible. The experimental approach under such conditions has little to recommend it. By contrast, in a moderately complex system, small alterations of the elements lead to relatively small perturbations of the remaining elements. There is a strong correlation between the system state before and after a minor change. In such a system, marginal tinkering and experimentation is far more feasible.

In addition, the path-dependence exhibited by complex systems can generate a second, almost opposite, limitation: an inability to perturb the system. Due to the possibility of becoming stuck as suboptimal local equilibria, marginal changes may be insufficient to dislodge a system from its current, undesirable state. How do complexity theorists propose to deal with the prospect of suboptimal equilibria that may require a more heavy-handed approach to reform? The answer to this question constitutes a second major theme in complexity approaches to policy.
Coping with Complexity

A Theory of Hayekian Interventionism

When large interventions are required to shift a complex system to a new equilibrium, we come up against the problem of predictability. Unless there is some way to quarantine the effects of changes in certain elements of the system, we can have little confidence that an intervention will make things better rather than worse. The idea of isolating effects of policy changes to subsections of the system points towards a second common theme in complexity-inspired policy theory: the idea of modularity (Simon; D’Agostino). Even if complex systems are made up of a large numbers of epistatically linked elements, there may be a way to partition the system into relatively self-contained subsystems, or modules. Such a system exhibits decomposability. For example, the many parts of a car are, obviously, quite interdependent. A larger engine requires a chassis that can bear greater weight, the size of seats is limited by the interior space of the cabin, and so on. Nevertheless, it is not necessary to design all the parts together. We can have a team that designs and assembles the radio and that does so quite independently of the team that designs and assembles the engine. Each team has a much simpler problem to solve than a team that needs to consider the vehicle as a whole, and can therefore proceed much more quickly to the optimal solution. As D’Agostino explains this approach, the trick “is not to attack such a problem head-on, but, rather, to divide it into parts, allocate the parts to teams, allow the teams to solve the resulting subproblems, and then assemble the solutions discovered by these teams to provide an overall solution to the problem in question” (2009, 109).

The modularity approach is by no means a panacea. Even in relatively decomposable systems, such as our automobile, interdependencies between the modules may prove important (Axelrod and Cohen, 107). Suppose, for example, that the ideal radio for the vehicle requires an electric current of 20 amps, but the ideal battery cannot produce this current — at least, not if we implement the proposed ideal headlights, wipers, wiring, GPS system, and vehicle weight. If the teams all go ahead and put their “ideal” modules together, the result will be a completely dysfunctional car. Importantly, this means that certain elements within each module set limitations on the elements within other modules (or vice versa).

Despite these challenges, if a system is relatively decomposable, interdependencies between its modules can be addressed by imposing design rules. Design rules limit the search space within modules by fixing the variables that determine functionality between modules. So, for example, if one of the variables that the radio team must decide upon is the number of amps required, a design rule may fix this variable at a reasonable 14, while fixing the corresponding variable in the car battery at some number greater than or equal to 14. If each team observes these design rules as they go about searching for the optimal design of their module, then when we assemble the various modules we will, at least, have a functioning vehicle. And the existence of this solution is important. For, as Marengo and Dosi put it, “problem solving by boundedly rational agents must necessarily proceed by decomposing any large, complex and intractable problem into smaller sub-problems which can be solved independently” (quoted in D’Agostino, 2009, 109).

A third theme echoed by many complexity theorists is the idea of addressing society as an organic, evolving entity, rather than a designed...
artifact (Colander and Kupers 2014, 55-8; D’Agostino 2009, 120-2; Gowdy et. al. 2016, 327; Axelrod and Cohen 2000, xvi and 155). Attempts to achieve precise results or execute a detailed plan by manipulating individuals within society ignores the fundamental fact that individuals are unpredictable and that they interact in unpredictable ways. Although society cannot be effectively controlled, as one might control a computer program, there may be ways in which we can positively influence society by altering the conditions in which it evolves. A common metaphor is to compare the role of the policy maker to that of the cultivator, in contrast to that of the engineer. Policy making, like the gardening, is most effective when it leaves the organism (society) free to grow as its internal principles direct it, exerting influence only on the environment in which that growth occurs. In gardening, this means paying attention to soil, fertilizer, water, and sun. In society, this means paying attention to the incentive structures that individuals face with an eye toward the effect of their actions on the whole order.

Perhaps the clearest statement of this idea, one which brings it beyond mere analogy, emerges from David Wilson’s distinction between two ways in which complex systems adapt and evolve: (1) CAS1, which refers to macro-level adaptations of the system as a whole, and (2) CAS2, which refers to the adaptive behaviors or strategies employed by the parts of the system (Wilson 2016, 31). Wilson, along with several co-authors, leverage this distinction in explaining the role of a policymaker when dealing with complex systems. Rather than disregarding CAS2 in an attempt to control CAS1 characteristics, the task is to align adaptation on both levels so that the adaptive behaviors of individuals, CAS2, actually lead to CAS1 adaptations, thus yielding benefits on the system level (Gowdy et. al. 2016, 328, 330, 340). When this alignment occurs, the independent and often myopic behaviors of various individuals serve a purpose beyond their immediate aims, contributing to a flourishing society that benefits all. In short, the task of the policy maker is to tweak the institutions and the incentives they offer so as to actualize Adam Smith’s proverbial invisible hand, in which each individual is “led... to promote an end which was no part of his intention” (Smith 1976, 184-5). Or, as expressed by two contemporary complexity theorists: “In the complexity policy frame, it is the result of a conscious attempt to develop an ecostructure without a central controller that is adequate to coordinate individuals’ actions” (Colander and Kupers 2014, 59).

**Hayekian Interventionism**

Our examination of themes in complexity theory has revealed that approaching policy with social complexity in mind tends to promote skepticism about the policymaker’s ability to control society without promoting skepticism about the policymaker’s ability to influence society. Hayek’s approach to interventions exhibits these same tendencies. In fact, his approach to intervention expresses many of the same central themes as contemporary complexity theorists.

(1) Consider, first, Hayek’s version of incrementalism. In his presentation of the idealized evolution of law and of social reform more generally, Hayek posits that changes to rules should be gradual. Large changes in the rules will disrupt the ability of individuals to develop reliable
expectations and thus to coordinate their actions. Accordingly, law should typically emerge from a judge whose main concern is not “what any authority wants done in a particular instance, but with what private persons have ‘legitimate’ reasons to expect, where ‘legitimate’ refers to the kind of expectations on which generally his actions in that society have been based” (Hayek 1983, 98). The judge, then, primarily seeks to understand an existing order so that she might determine what the most reasonable set of expectations would be within this order. In this sense, the judge is often not a creator of rules, but merely a student and an articulator of pre-existing rules, which she formulates as explicit legal precedents. When situations arise where there is no pre-existing rule that would determine whether one set of expectations is more reasonable than another, the judge is tasked with creating a new one. In such a situation, however, the change to the rules underlying the spontaneous order is relatively minor — it aims merely to fill in a gap in the existing rules and to do so in the least disruptive manner possible (ibid., 101). The role of the judge, as Hayek’s preferred source of laws and legal reform, is that of “moving within an existing system of thought” and to employ a method “of piecemeal tinkering” (ibid., 118). The idea of piecemeal tinkering also appears in Hayek’s account of moral reform. For example, in “The Errors of Constructivism,” Hayek outlines what he takes to be a non-conservative but still responsible approach to moral reform:

The proper conclusion from the considerations I have advanced is by no means that we may confidently accept all the old and traditional values. Nor even that there are any values or moral principles, which science may not occasionally question. The social scientist who endeavours to understand how society functions, and to discover where it can be improved, must claim the right critically to examine, and even to judge, every single value of our society. The consequence of what I have said is merely that we can never at one and the same time question all its values. ([1970] 2014, 352)

By affirming or rejecting values in isolation, and doing so only in light of their compatibility with the rest of our values, we maintain the integrity of our current order and avoid drastic changes yielding unpredictable results.20

(2) Hayek assigns a special term, immanent criticism, to the method of evaluating particular rules or values only in light of their coherence within the the total system of rules and values. The purpose of immanent criticism is “to make the whole more consistent both internally as well as with the facts to which the rules are applied” (Hayek 1983, 118), and in this capacity, immanent criticism points toward the way in which Hayek’s interventionism exhibits a second major theme of complexity-based policy: modularity. Immanent criticism suggests that when altering rules, we do not simply intuit what rule would be best, nor do we apply a simple criteria, based on a narrow set of concerns (as primitive versions of act utilitarianism would recommend). Rather, we take a system-level perspective and consider the connections between the rules we wish to modify and the other rules within the system. In complex systems, large changes yield unpredictable results, but as we have seen in discussing D’Agostino’s work on modularity, if systems are relatively decomposable, large changes can be made within subsystems so long as design rules are

---

20: Importantly, “values” is a term that Hayek uses very broadly. It refers also to what we would call rules, rights, and duties. See (Hayek [1970] 2014, 343).
in place to maintain compatibility between these subsystems. Does this idea find expression in Hayek’s writings on intervention?

In Hayek’s view, institutions and rules are at least somewhat decomposable in the sense explained above: there are subsets of rules that exhibit strong interdependencies among one another and weaker connections to elements in other subsystems. As Hayek characterizes the Great Society, it is comparable to “a nucleus, or several nuclei, of more closely related individuals occupying a central position in a more loosely connected but more extensive order” (1983, 47). These loose partitions are made up of distinct, but overlapping subsets of individuals, and are also governed by different sets of rules. For example, the rules of a military are distinct from the rules of citizens, but because individuals are members of both categories, these sets of rules are not entirely independent.

This version of modularity provides a clue for making sense of Hayek’s non-incremental interventions. For instance, when Hayek discusses non-incremental changes in the law, undertaken by legislators, he is careful to emphasize the system-level perspective. These changes are desirable only in considering the incoherence between subsets of rules and the undesirable “dead end” to which inevitably path-dependent rule evolution can lead. Just as D’Agostino points out about design rules, when functionality considerations lead us to fix rules that connect two subsystems, this may create a path dependency by restricting searches to sets that satisfy this design rule. The law will not evolve in such a way that could alter this design rule, since doing so in an effective manner would require a systematic or non-incremental alteration in the rules. As Hayek explains, “The development of case-law is in some respects a sort of one-way street: when it has already moved a considerable distance in one direction, it often cannot retrace its steps when some implications of earlier decisions are seen to be clearly undesirable” (ibid., 88). Legislation, when the suboptimal equilibrium is sufficiently obvious, can correct these path dependent lock-ins, because, unlike judge-made law, legislation is not confined to piecemeal tinkering. Notably, Hayek explicitly mentions that lawmakers should strive to understand the properties and requirements of the system as a whole and that this systemic perspective should influence their judgments; lawmakers must decide cases “in a manner appropriate to the function which the whole system of rules serves” (1983, 116; see also Mack 2006, 279-80).

Consider a difficult example for Hayek’s conception of law: slavery. If the institution of slavery becomes entrenched in a society, the gradual evolution of judge-made law is unlikely to dislodge it. Too many expectations depend upon the persistence of this institution. In many countries, though not the United States, legislative intervention succeeded in abolishing the institution while minimizing the disappointment of expectations. The rules surrounding the ownership of slaves can be seen as a somewhat modular subsection of property law, which itself is a subsection of law in general, with important connections to other modules within the system of rules, including moral rules. As moral rules and values evolved, slavery became increasingly incompatible with common opinion regarding basic rights and interpersonal rules of conduct. Since laws emerge together, however, there existed tight connections between various laws (Hayek 1983, 65); case law had produced a lock-in effect, or path dependency, which was impossible to undo through piecemeal tinkering. Too
many interests depended on the persistence of slavery. Too many plans had been laid on the assumption of its continued existence. And too many other laws (and legal precedents) depended on and supported the institution of slavery. Instead of piecemeal tinkering, a less gradualist intervention, one based on immanent criticism, was required. Thus, when Hayek advocates “the revision not only of single rules but of whole sections of the established system of case law,” he is best understood as advocating a form of system-level or modularity thinking in which the compatibility between various subsets of values drives moral and legal reform. This focus on immanent criticism explains why, when discussing the creation of new laws, Hayek often points out that these reforms do not occur in a vacuum, but rather that “hereto accepted rules are unjust in the light of more general principles of justice” (1983, 89) — i.e. in light of other values and rules within the system or within connected systems. The legislator, just like the judge when he must create new rules, is focused on the system as a whole and on the order that it generates. In doing so...

... the only standard by which we can judge particular values of our society is the entire body of other values of that same society. More precisely, the factually existing, but always imperfect, order of actions produced by obedience to these values provides the touchstone for evaluation. (Hayek [1970] 2014, 354)

(3) As for the third theme — that of cultivating, rather than controlling — Hayek advances a vision of intervention in perfect alignment with the viewpoint of modern complexity theorists. Just as contemporary complexity theorists criticize policy approaches that fail to take account of social complexity, ascribing labels such as “state control policy” (Colander and Kupers 2014, 44) or pointing out an undue focus on “allocation” rather than “formation” (Gowdy et. al. 2016, 328), Hayek also criticizes more common ways of understanding the role of policy as a means of control ([1955B]2014, 163). Because we possess only knowledge of the principles of how complex systems operate, and are thus limited to mere “pattern predictions,” policymakers cannot predict, and thus cannot control, a complex system like society (Hayek [1975] 2014, 365). Nevertheless, the policymaker may be in a position to support the ability of individuals to coordinate with one another and to pursue their respective ends. As Hayek explains in “Degrees of Explanation”:

... Even if we cannot control the external circumstances at all, we may adapt our actions to them. And sometimes, though we may not be able to bring about the particular results we would like, knowledge of the principle of the thing will enable us to make circumstances more favourable to the kinds of events we desire... An explanation of the principle will thus often enable us to create such favourable circumstances even if it does not allow us to control the outcome. Such activities in which we are guided by a knowledge merely of the principle of the thing should perhaps better be described by the term cultivation than by the familiar term ‘control’ — cultivation in the sense in which the farmer or gardener cultivates his plants, where he knows and can control only some of the determining circumstances, and in which the
wise legislator or statesman will probably attempt to cultivate rather than to control the forces of the social process. ([1955A] 2014, 210)33

The main method that Hayek proposes for cultivating a successful society is to promote general, equal, and predictable laws, or what Hayek labels the Rule of Law. With such laws in place, the spontaneous order is likely to emerge from the interactions between individuals. The requirements of generality, equality, and certainty still leave ample room for variety. The policymaker should aim, within these constraints, to determine which policies best promote the capacity of individuals to interact in beneficial and predictable ways.

To borrow Wilson’s terms, Hayek believes that the task of the policy maker is to set up conditions under which the behavioral adaptations of individuals, the CAS2 adaptations, promote positive change on the social level, i.e. generate positive CASI adaptations. But this requires understanding the nature and conditions of individuals’ self-organizing capacities: “our main task must be to adjust our rules so as to make the spontaneous forces of society work as beneficially as possible. The first need in order that we should be able to do so is that we learn to understand the working of those forces” (Hayek [1955B] 2014, 192).

Hayek’s comments on intervention therefore bear a striking resemblance to more contemporary views that approach policymaking from a complexity perspective. This resemblance appears in various forms; I have emphasized three key themes that characterize Hayek’s viewpoint as well as that of more contemporary theorists. From a complexity perspective, incremental changes are typically better than large-scale ones. Yet, path dependency and lock-in may demand exceptions, and when exceptions must be made, complexity requires system-level thinking. Accordingly, policymakers must pay attention to any potential for decomposability within the system. Otherwise large-scale interventions are likely to severely disrupt the social order and to unleash a host of unintended consequences. Finally, the approach recommended by the complexity perspective is that of cultivating a successfully evolving order, of reconciling adaptations on the CASI and CAS2 levels, rather than attempting to control CASI characteristics directly. I have already suggested how complexity considerations may shed light on Hayek’s seemingly oxymoronic position on intervention. Before concluding, the ability of complexity theory to reconcile Hayek’s claims about intervention must be made more explicit.

### 2.6 Addressing the Puzzle

Having seen that Hayek’s theory of complexity sheds light on both his opposition to and advocacy of interventionist measures, it remains to be seen if this new understanding can also aid in reconciling these two strands in Hayek’s writings. Recall the two claims presented in section III that give rise to a perceived tension in Hayek’s stance on intervention:

1. Interventions cause great harm
2. Interventions of various sorts are compatible with or necessary for a healthy society
Taking (I) as given, how can Hayek endorse (2)? To examine this issue, consider two different types of intervention that Hayek discusses: governmental programs and legal or institutional reform. Reading Hayek as a complexity theorist provides the key to resolving the tension that arises with regard to both types of intervention.

First consider particular governmental programs, such as a social safety net or the provision of public goods. Whether or not such programs create harmful disturbances in a complex system may depend on how they are implemented and on how they interact with currently existing laws and norms. Hayek has relatively little to say about the implementation of these programs, but his comments on the importance of piecemeal tinkering as well as immanent criticism suggest that a Hayekian intervention would refrain from imposing large-scale programs or regulations which would be both disruptive and unpredictable. Instead, small changes or small-scale implementation better fit the Hayekian vision of policy aimed at facilitating reliable expectations and interpersonal coordination under conditions of complexity.

Even with this approach to implementation, however, certain policies are simply off the table for Hayek. As Hayek explains in his Cairo lectures, “all laws and institutions which offend against the ideal of the Rule of Law are objectionable in principle, while any law which conforms to it will have to be judged on its individual merits” (ibid., 178). A necessary condition for any policy or rule change is that it satisfy three conditions of the Rule of Law: generality, equality, and certainty (ibid., 172). Although there are many ways in which a law or policy can be harmful, if interventionist activities are restricted by generality, equality, and certainty, then the individual retains the ability to respond to a reasonably predictable environment, one insulated from the whims of authority (ibid., 178).

The thought is that, given the unpredictability of complex systems, restricting laws in this manner is necessary in order to “make the world around us a more familiar world in which we can move with greater confidence that we shall not be disappointed because we can at least exclude certain eventualities” (Hayek [1955A] 2014, 209-10). Admittedly, this account places a heavy burden on a somewhat dubious distinction — interferences that satisfy the Rule of Law and those that do not. Hayek claims that price-fixing, for example, cannot satisfy the Rule of Law ([1955B] 2014, 180). Yet, price-fixing might be applied in a manner that is general (applying to all cases that are relevantly similar), equal (applying equally to all citizens), and certain (predictably applied). Given that the policy implications of complexity theory are still being explored and debated, it is perhaps unsurprising that Hayek’s early treatment falls short of perfection. Still, it would be rather scandalous if Hayek’s theory lacked the resources to distinguish between the interventions he entertains and those he claims are simply out of the question.

A more helpful distinction than that between laws that do and laws that do not satisfy the Rule of Law is the distinction between principle and expedience. Although this distinction appears briefly in the Cairo lectures (ibid., 180), Hayek develops it further in volume one of Law, Legislation, and Liberty. Principles are inchoate rules, delimiting which concerns do and which do not permit a change in the laws (Hayek 1983, 55). Their purpose is to supersede case-by-case assessment of particular
actions or policies, which would lead to myopic decision-making, taken in the heat of the moment.\textsuperscript{23} Expedience, on the other hand, focuses, not on the sorts of considerations which may justify a rule change, but on the desired outcome to be achieved (Hayek [1955B] 2014, 180). In Hayek’s view, principles lead to predictable “interventions,” consistent with the functioning of a spontaneous order in complex systems, while decisions based on expediency are unpredictable and thus undermine the formation of spontaneous order. Recalling the third theme of the complexity approach, we see that expedience, in virtue of its specific goals, leads to attempts at control, rather than cultivation. Principle-based policy, on the other hand, lacks specific purposes for particular persons, seeking instead to facilitate an “order of actions,” which promotes the ability of individuals to pursue their own diverse plans.

This distinction clarifies how many economic controls, like price-fixing, fall outside the realm of acceptable interventions. If we are constrained to apply rules only in light of principles, not considering immediate expediency, then price controls lack appeal. For why would one endorse a policy of price control if not to benefit a particular group at a particular time, i.e., on the basis of expediency?\textsuperscript{37} A price control may help a select group of sellers, but it does so at the expense of other prospective sellers, of consumers, and ultimately of economic efficiency.\textsuperscript{38} The only reason to institute such a policy is to achieve a particular purpose at a particular time — but, as we have seen, such endeavors are incompatible with facing the challenges of social complexity, since they undermine the expectations of interacting individuals. While ruling out many of the interventions that Hayek opposes, this principle-expediency distinction seems to allow for many toward which Hayek is more ambivalent: for example, a social safety net and the provision or regulation of public goods. As Hayek recognized, exactly which interventions are or are not justified by principles is debatable. Where Hayek is confident is in his assertion that when evaluating the acceptability of a policy on the basis of a principle, complexity considerations play a central role. In order to function under conditions of complexity, society must facilitate accurate expectations and interpersonal coordination. It cannot do so when expediency, rather than principle, drives policy.

Hayek’s theory of complexity also suggests a reconciliation between his favoring of spontaneous order and direct legislative intervention. A crucial point, which has been implicit, is that a spontaneous order is distinct and separable from the spontaneous origin of the laws that give rise to it: “... while the rules on which a spontaneous order rests may also be of spontaneous origin, this need not always be the case... it is at least conceivable that the formation of a spontaneous order relies entirely on rules that were deliberately made” (Hayek 1983, 45).\textsuperscript{39} This is why Hayek need not rule out legislative intervention: rules are not good or bad in virtue of their source, but rather, in virtue of their form. Do they or do they not allow us to cope with complexity by enabling individuals to develop accurate expectations and to succeed in coordinating their activities? Importantly, Hayek does not say that laws or legislation are incompatible with spontaneous order; he says that commands are:

\textit{It is advisable... not to confuse laws and commands... The important difference between the two concepts lies in the fact that, as we move from commands to laws, the source of}
the decision on what particular action is to be taken shifts progressively from the issuer of the command or law to the acting person. (Hayek 2011, 218)

While commands are appropriate for organizations, in which members are directed by a central authority and work towards a single goal, they cannot engender a spontaneous order of actions, in which individuals pursue their own ends and are coordinated by formal rules that do not specify particular ends. The legislator who finds it necessary to correct an evolutionary “dead-end” in case law need not issue commands and therefore need not undermine the spontaneous order or mould society as a whole into the form of an organization. Instead, the legislator can restrict new rules to the form of laws. Combined with the complexity considerations covered in section V — specifically, the system-level thinking of immanent criticism and the potential for focusing on modules within the system of rules — this distinction between commands and laws reveals how legislative reform of “whole sections of the established system of case law” is less drastic and disruptive than it sounds. By isolating these “sections” into relatively self-contained modules while also focusing on important connections these modules have to other sections of the system of law, the legislator can maintain or even enhance the ability of individuals to from expectations and to coordinate, at least so long as the legislator issues bona fide laws and avoids the temptation to issue commands.

On the flipside, it is precisely due to complexity that such heavy-handed interventions sometimes become necessary. Complex systems are characterized by positive feedbacks and path dependence (Arthur 2013, 13-4; Marengo and Dosi 2013). For this reason, Hayek explains, “spontaneous process of growth may lead into an impasse from which it cannot extricate itself by its own forces or which it will at least not correct quickly enough” (Hayek 1983, 88; see also Vaughn 1999, 249). In correcting such undesirable directions of growth, a prevailing concern is to maintain society’s ability to spontaneously order itself; policy is to be “directed toward the securing of an abstract overall order” (Hayek 1978, 114). Hayek’s concern with this order of actions, so crucial to interpersonal coordination in a complex system, again motivates his emphasis on principle over expediency. Aiming at particular outcomes, as expediency dictates, would yield policies and rules incompatible with the requirements and limitations presented by complex systems. Laws based on principle, on the other hand, are better suited to decentralize decision-making to dispersed actors, thereby supporting the spontaneous forces that give rise to the order of actions.

2.7 Conclusion

Hayek the laissez-faire marketeer is an urban legend. Yet, his writings express a strong anti-interventionist streak. Concerns with diversity, with knowledge, and with predictability lead him to conclude that interventions are often harmful and unjustified. To focus only on this aspect of his position, however, is to mistakenly view his stance on intervention as one of unequivocal opposition. In fact, there are many interventions that Hayek believes are worth considering, interventions
which cannot be ruled out on principle. To understand Hayek’s views on intervention requires that we search out exactly what divides those interventions that one can rule out on principle from those that must stand or fall on the basis of a consideration of their particular merits. To this end, I have suggested that we pay closer attention to Hayek’s theory of complexity and how his understanding of society as a complex system underlies his stance on intervention. Complexity provides insight into why certain interventions — those that do not disrupt spontaneous order — are permissible. Whether or not Hayek’s complexity approach to intervention is fully satisfactory, considering his position in light of complexity certainly contributes to our understanding of statements that otherwise appear capricious or contradictory. The restrictions Hayek places on intervention closely track restrictions on successful intervention into a complex society. Those he allows are those that maintain our capacity to cope with complexity.
3.1 Introduction

In one breath F. A. Hayek will urge the primacy of “orderly structures which are the product of the action of many men but are not the result of human design” (1973, p 37); in the next, however, he will emphasize that social structures owe their persistence or extinction to their tendency to “increase or decrease the efficiency of the group as a whole” ([1967A] 2014, p 283). Critics have claimed that Hayek thus posits two distinct accounts of the emergence of social institutions: one which emphasizes the local adjustments of individuals unintentionally giving rise to social structures, another which emphasizes the effectiveness of macro-level properties and their adaptiveness in the context of intergroup competition. “Hayek’s appeal to a process of cultural evolution operating at the group-level as such,” writes Viktor Vanberg, “stands in contrast to his explicit methodological individualism. . . and to what is otherwise the main thrust of his work” (1986, p 85). This inconsistency, according to Vanberg, “represents a major flaw in Hayek’s reasoning” (1986, p 85).

The tension between these two processes is not an idiosyncratic error in Hayek’s system of thought. Rather, it is endemic to the dynamics of group adaptation more generally. Theories of multilevel selection acknowledge that adaptation occurs on a multiplicity of scales. A simple, illustrative model of multilevel selection posits two levels: individual organism and social group. Based on these disparate levels of selection, the evolutionary biologist David Sloan Wilson defines two types of adaptive system:

- Complex Adaptive System 1 (CAS1): “A complex system that is adaptive as a system.”

CAS1 adaptations occur when systems or groups adjust to their environments, environments which include other systems or groups. Group selection is an example of a process of CAS1 adaptation. By contrast, CAS2 adaptations occur via the local adjustments made by individual elements that make up a system or group. Invisible hand processes, in which individuals pursue values and unintentionally produce macrostates, are paradigmatic of CAS2 adaptation. Importantly, CAS1 processes will involve changes in individual behavior, just as CAS2 processes will generate group-level modifications.

The distinct nature, and alleged incompatibility, of these two processes has led to a dialectic in which evolutionary theorists — be they biological or cultural — tend to champion one of two broad camps: those who grant primacy to selection at the group level and those who emphasize invisible hand processes that ultimately give rise to group characteristics (Mesoudi 1999, 19-21).
3 Hayek’s Twin Ideas: Reconciling Methodological Individualism and Group Selection

2011, pp 51-3). For conciseness, the former group will be referred to as “CAS1 theorists,” the latter as “CAS2 theorists.”

Much of the literature that examines Hayek’s use of evolutionary theory asks whether he is (or should be) a CAS1 or a CAS2 theorist. Whichever answer he provides, commentators claim, will entail serious alterations to his theory. If Hayek grants primacy to intergroup competition, they claim, then the role of individual-level adaptations is diminished. Hayek is no longer a bona fide methodological individualist, seeking to explain culture and institutions as the results of “human action but not human design.” Not only does this rejection of methodological individualism put Hayek at odds with his earlier work, which champions the “compositive method” of deducing macro-results from individuals’ behavior ([1952B] 2010; [1943] 1980), but it also calls into question Hayek’s self-identification as a philosophical descendant of the Scottish Enlightenment. Invisible hand explanations feature prominently in Hayek’s reading of theorists such as Adam Smith and Adam Ferguson, whose philosophical movement Hayek takes himself to be reviving and extending ([1967A] 2014, pp 96 ff.).

On the other hand, if Hayek abandons his group-selectionist bent (CAS1) and wholly embraces methodological individualism (CAS2), then he faces the challenge of explaining why the emergent properties would be desirable or beneficial. As Jonny Anomaly and Geoffrey Brennan (2014) explain, for any invisible hand process that results in a benefit to the group, there also looms an invisible fist process that threatens to undermine group functionality. Wilson’s view is even more pessimistic: not only are individual adaptations logically unrelated to group level fitness, they actually tend to undermine it. In many cases, opportunistic adaptations on the individual level can engender “a breakdown of group-level functional organization,” because “[b]etween-group selection, if it takes place at all, is not sufficiently strong to oppose within-group selection. The groups remain functionally impaired indefinitely...” (Wilson 2016, p 38). If Hayek emphasizes local adjustments over group adaptations, therefore, he weakens the social-scientific basis of his anti-interventionist political stance. For, if CAS2 adaptations cannot be expected to generate socially beneficial outcomes, then the government must actively endeavor to make individual behavior compatible with group functionality (Wilson 2016, p 45).

In sum, several commentators agree that Hayek must either abandon methodological individualism as the explanatory paradigm of social science or he must abandon his group-selectionist arguments. And they also agree that neither course leaves his broader theory intact.

This supposed dilemma achieves plausibility only by mischaracterizing Hayek’s view. Far from conflating two distinct evolutionary processes, Hayek incorporates both processes as integral features of his broader theory. Hayek’s view is best interpreted as a multilevel selection theory involving both upward and downward causation. The argument for this interpretation of Hayek’s view unfolds as follows: section II explains why neither group selection nor methodological individualism can provide an adequate explanation of social order when considered in isolation. By combining these two approaches, section III argues, Hayek provides a more satisfying explanation. Pace Hayek’s critics, these two fundamental

3: Daniel Klein (1997) has noted that Menger himself thought there was little reason to expect evolved conventions to yield efficient outcomes.

4: Although, as Jerry Gaus pointed out to me, an unstated assumption here is that a central governor would outperform CAS2 processes.
approaches hang together as a unified framework for explaining social phenomena.

### 3.2 The Inadequacy of CAS1 and CAS2 Approaches

Hayek’s critics make two assertions. First, in advocating for a CAS1 or CAS2 approach, they suggest that Hayek’s evolutionary theory or his methodological individualism, respectively, provide an adequate explanation of social order. Second, they hold that this “or” is exclusive: Hayek’s theory of group selection and his methodological individualism are incompatible accounts.

Before assessing CAS1 and CAS2 approaches as stand-alone explanations, it is necessary to briefly consider what makes for an adequate explanation. Hayek often cites two philosophers of science, Karl Popper and Ernest Nagel, whose theories are representative of the prevailing philosophy of science in Hayek’s day. Both of these philosophers advocated versions of the deductive-nomological theory of scientific explanation, in which explanations consist of a set of propositions, at least one of which is a scientific law, that jointly necessitate the outcome to be explained. In a book dedicated to Hayek, Popper explains that, for a theory to constitute an explanation of observable facts, “a description of these facts must be deducible from the theory in conjunction with certain statements, the so-called ‘initial conditions’” (Popper 1968 [1962], p 103). This sets a rather high bar for Hayek: an explanation of social order must include empirical and law-like propositions that necessitate the emergence of that order. To clear this high bar, I will argue, Hayek must employ both types of explanation: group selectionist and methodological individualist.

#### Downward Causation: Hayek’s Evolutionary Theory

According to Vanberg, CAS1 theorists advocate “collective functionalism,” a view which posits a “direct link” between benefit to the group and the emergence of a macrostructure (1986, pp 83–4). Such explanations clash with methodological individualism due to the fact that individual advantage and group benefit often pull in opposing directions. For Vanberg, this means that the predictions of a collective functionalist view are implausible, since they fail to account for the decisions made by individuals on the basis of their personal interests. Worse yet, collective functionalist explanations are also vague, because the mechanism by which collective functionality actually gives rise to macrostructures remains unspecified. Given that Hayek’s theory of cultural evolution is a type of collective functionalism, Vanberg asserts, it thus exhibits both major issues: implausibility and vagueness. These issues set the challenge for Hayek’s use of group selection. Can a CAS1 approach avoid implausibility by explaining how self-interested behavior is compatible with group interest? And can it avoid vagueness by explaining the mechanism by which group benefits cause group adaptations? Without resolving both of these issues, CAS1 theory cannot provide a bona fide
explanation, at least not according to the deductive-nomological theory outlined above.

With respect to the challenge of implausibility, contemporary group selection theory has made major strides in clarifying how group selection might occur in the face of individual-level selective pressures. Contemporary group selectionists, such as Sober and Wilson (1998), have clarified how group-level pressures can overcome individual-level ones, even when these pressures direct behavior in opposite directions. Multilevel selection theorists often point to Price’s equation to clarify the conditions under which pressures at a higher level will swamp those at a lower level in the evolutionary process (Turchin 2011, p 8; Okasha 2009, pp 62 ff.). When within-group diversity is small, between-group diversity is large, and when individuals within groups have relatively equal fitness, group selection can successfully overpower individual-level selective pressures.

When group selection occurs, the result is what evolutionary theorists call “downward causation” (Flack 2017; Corning and Szathmáry 2015; Ellis 2012; [1967A] 2014, p 288). Lower-level elements tend to exhibit those traits and behaviors that prove to be adaptive for the group. At the limit, the group can actually become a “superorganism,” programming its constituent individuals to behave in coordinated, group-beneficial ways (Szathmáry 2015). This process has been thoroughly explored by models that demonstrate its possibility under various circumstances. It has also been verified empirically (Wade 2003; Maynard-Smith and Szathmáry 1995; Lewontin 1970; Hamilton 1967). There is therefore no justification, at least given what we know today, for Vanberg’s assertion that Hayek’s theory of group selection is “implausible.” Indeed, in offering his own criticism of Hayek, Geoffrey Hodgson makes the same point: “Vanberg was wrong to dismiss group selection so easily. . . . [A]mong biologists the notion of group selection is not so reprehensible as Vanberg suggests, and it retains some considerable support from a number of theorists and philosophers” (1991, p 69).

There is another criticism of group selection theory, different from Vanberg’s, that also claims to show its implausibility. So-called “gene reductionists” claim that groups are not suitable as units of selection (Williams 1996). As Richard Dawkins memorably put it,

[g]enetically speaking. . . groups are like clouds in the sky or dust-storms in the desert. They are temporary aggregations or federations. They are not stable through evolutionary time. . . . A population is not a discrete enough entity to be a unit of natural selection, not stable and unitary enough to be ‘selected’ in preference to another population. (1989 [1976], p 34)

Groups lack the longevity and stability of genes. Hence, they cannot function as “replicators,” i.e. as units of selection.

Although critics of group selection are right to attack naïve accounts that take groups to be replicators, this is not Hayek’s view. For him, the replicator is a system of rules, not the group or even the emergent order of actions: “the natural selection of rules will operate on the bases of the greater or lesser efficiency of the resulting order of the group” ([1967A]
Hayek’s Twin Ideas: Reconciling Methodological Individualism and Group Selection

Selection takes place at the level of social order, the “vehicle” (Dawkins 1989 [1976], p 19), but the thing selected, the “replicator,” is the system of rules.

Having identified replicators as rule systems and vehicles as social orders, Hayek is able to sidestep the criticism that groups are too unstable to be replicators. They aren’t the replicators, they are the vehicles. In this way, Hayek was ahead of his time. As Kim Sterelny explains, recent theories of group selection are best understood as claiming that “groups function in evolution in the same way organisms do. That is, group selectionist hypotheses are about vehicles, not replicators” (Sterelny, p 564). Much of Hayek’s work on cultural evolution aims to demonstrate that systems of rules exhibit heredity, variation, and selection — sufficient conditions for evolution to take place (Mesoudi 2011, pp 27-34).

Hayek’s CAS approach stands up well to the criticisms of both Vanberg and Dawkins that call into question the plausibility of group selectionist accounts. However, it remains to be seen whether or not Hayek’s account is guilty of vagueness, as Vanberg asserts. All theories, to some extent, stylize, so vagueness falls along a spectrum. However, recalling Hayek’s commitment to the deductive-nomological theory of scientific explanation, we might posit that to be “too vague” for Hayek means that a theory fails to provide a set of sufficient conditions entailing the explanandum. Taking this view, it’s clear that Vanberg is correct: there is a major gap in Hayek’s group-selectionist theory of social order. Although Hayek adequately spells out the process of downward causation from group-beneficial order to the proliferation and extinction of certain systems of rules, group selection alone fails to specify how systems of rules actually produce order. Some rules will generate beneficial orders, some will generate dysfunctional orders, and some will fail to produce order at all (1973, p 44). Without explaining how this process works, Hayek’s social theory does not meet the requirements of scientific explanation, since there is a largely unexplained link between certain systems of rules and certain group-level traits. In other words, despite providing a detailed account of downward causation, group selectionism says nothing about the process of upward causation.

Upward Causation: Methodological Individualism and Spontaneous Order

In Vanberg’s view, methodological individualism and spontaneous order, closely related concepts, provide the main alternative model for explaining social phenomena. For Hodgson, however, this explanatory method fails to accomplish Hayek’s goals. Hodgson criticizes methodological individualism, first, for being incomplete (Hodgson 1991, pp 78-9). Methodological individualism, as Hodgson understands it, takes individual values for granted, eschewing explorations into psychology and preference-formation. From there, it attempts to work up explanations of human behavior by individuals acting within groups. By leaving the origins of human nature unexamined, Hodgson asserts, Hayek fails to provide a full account of “long-run changes and developments in a social context” (78). An adequate explanation of social order must therefore “involve a discussion of the molding of individuals by circumstances
as well as of circumstances by individuals,” that is, it must provide an “evolutionary perspective” (79). Hodgson also questions whether Hayek’s social theory is consistent with his anti-interventionist stance (80). This second criticism is echoed most clearly by Wilson:

... there is an unstated assumption that [CAS1 and CAS2] are compatible with each other; in other words, that a system composed of agents which employ adaptive strategies is adaptive as a system. The metaphor of the invisible hand, which has played such a central role in the history of economic thought, embodies the same assumption. But a proper understanding of evolutionary theory leads unequivocally to the conclusion that these two meanings are not compatible with each other, except under special conditions (2016, p 33).7

According to both Hodgson and Wilson, there is no reason whatsoever to believe that individual-level adaptations will facilitate collective functionality.8 In fact, given the prevalence of social dilemmas, individual-level adaptations are likely to undermine group success. Wilson, associating Hayek with the view that “greed is good,” thus finds Hayek’s use of evolutionary theory ironic (Wilson 2015, pp 95-6, 100).

Examining the veracity of this criticism requires at least a rough understanding of how Hayek understands methodological individualism and how he employs it to explain spontaneous social orders. Both the meaning of “methodological individualism” and its relation to the idea of spontaneous order must be examined.

For Hayek, the basic idea of methodological individualism is that an explanation or understanding of social phenomena must begin with an analysis of individual human action: “there is no other way toward an understanding of social phenomena but through our understanding of individual actions” ([1952C] 2010, p 6). This does not mean the theorist cannot speak of groups or of collective action. But such concepts must, ultimately, be analyzed in terms of the behavior of individual actors.9 The need for an analysis of individual human action comprises at least two features. The first feature, often referred to as Verstehen, concerns the need for interpretive understanding in the explanation of human behavior.10 Hayek concisely characterizes this methodological postulate by saying that “[s]o far as human action is concerned, things are what the acting people think they are” ([1943] 1980, p 89).11 As an important corollary of this truism, costs and benefits must be construed subjectively, and human action, therefore, cannot be understood until one comprehends the agent’s subjective viewpoint.

Equally important is Hayek’s explanation of why social science must begin with an analysis of individual human action. According to Hayek, social phenomena are often too complex or too abstract to be perceived without a theory that constructs them from the behavior of interacting individuals. Consequently, methodological individualism is not merely a tool of explanation, but also a tool of discovery. In the social sciences, Hayek claims,

... it is the attitudes of individuals which are the familiar elements and by the combination of which we try to reproduce

7: Note that Wilson does apply his critique directly to Hayek’s work in his book Does Altruism Exist? (2015, pp 95-6). His criticism is marred by the fact that he ignores the importance that Hayek places on the institutional framework that channels CAS2 processes. See Boettke (2019, pp 25).
8: In fact, many spontaneous orders, produced by individuals adhering to local norms, are rather horrific; they are far worse than close alternatives or even than having no norm whatsoever governing that specific domain. Henrich (2017), offers a host of gruesome examples of suboptimal norms that would qualify as spontaneous orders. For a cringe, see page 161.
9: Caldwell (2004, pp 279-87) identifies this as one of the areas in which Hayek’s conception of methodological individualism overlaps with the more common mainstream notion of methodological individualism. At the same time, he is adamant that Hayek rejects (much of) the version of methodological individualism that posits an idealized model of rational economic man.
10: Denis (2014) emphasizes Verstehen in his account of methodological individualism.
11: This methodological postulate traces back through Ludwig von Mises, all the way to Carl Menger. Mises writes, “It is illusory to believe that it is possible to visualize collective wholes. They are never visible; their cognition is always the outcome of the understanding of the meaning which acting men attribute to their acts. We can see a crowd, i.e., a multitude of people. Whether this crowd is a mere gathering or a mass (in the sense in which this term is used in contemporary psychology) or an organized body or any other kind of social entity is a question which can only be answered by understanding the meaning which they themselves attach to their presence. And this meaning is always the meaning of individuals. Not our senses, but understanding, a mental process, makes us recognize social entities” (2007 [1949], 43).
the complex phenomena, the results of individual actions, which are much less known — a procedure which often leads to the discovery of principles of structural coherence of the complex phenomena which had not been (and perhaps could not be) established by direct observation. ([1952B] 2010, pp 101-2)12

These complex phenomena — for example, the nature of general equilibrium in a market economy — are abstract, not concrete, and therefore not available to perceptual observation.13 Consequently, the social scientist cannot proceed by analyzing some perceivable whole into its basic elements, as Hayek believed was the common approach in the physical sciences. Instead, when explaining phenomena of high complexity, the social scientist must start with the basic elements and reconstruct these complex phenomena before such phenomena can be understood or sometimes even recognized.

Does Hayek mean to assert that we can deduce complex phenomena by identifying the basic elements and processes that give rise to them? There is good reason to reject this position whether or not Hayek held it. First, a hallmark of emergent phenomena is that one cannot infer their properties from the underlying elements; emergent phenomena are, in some sense, surprising.14 Many of the complex wholes that Hayek seeks to understand are emergent phenomena in this sense. Moreover, they are complex. The mind-boggling combinatorics involved in complex systems — systems with large numbers of interrelated, heterogeneous parts — pose serious challenges for any attempt to reverse engineer them (Green 2015). Finally, Hayek himself rejects the ability of social science to explain or predict particular outcomes in complex systems, like society.15 We cannot, therefore, trace out the implications of individual actions that give rise to particular macrostates.

Hayek does, however, seem to hold that social science enables us to understand certain relations or patterns exhibited by complex or abstract systems. Hayek calls these results “pattern predictions” and explanations ([1975] 2014, p 371). We cannot explain or predict the exact list of prices at a given moment in time; we cannot explain or predict the exact phenotype of a particular species at a given moment in the future. However, in both cases, we can start with simple assumptions about how basic elements behave and, from there, trace out a process that will reveal general patterns. Markets will tend toward market-clearing prices; genetic evolution will tend toward organisms that are well adapted to their environments, which include other organisms. In both cases, we rely on basic constituents to reveal general patterns that will emerge on the macro-level. The intellectual task of the social scientist is thus poorly described by the term analysis; it is, instead, a synthesis or composition ([1952B] 2010, pp 101-2). Accordingly, the term “explanation by methodological individualism” will be used interchangeably with the more concise “compositive explanation.”

This expanded construal of Hayek’s methodology captures the intuitive notion of methodological individualism that critics ascribe to Hayek. In particular, the core of this more detailed conception of methodological individualism is the reconstruction of emergent social phenomena — such as spontaneous orders — in terms of individual behavior. It thus

12: See also ([1943] 1980: 98), ([1967B] 2014: 70), ([1964] 2014: 259) and ([1943] 1980: 89-90): “Social theory. . . is, then, logically prior to history. It explains the terms which history must use. . . The social complexes, the social wholes which the historian discusses, are never found ready given as are the persistent structures in the organic (animal or vegetable) world. They are created by him by a work of construction or interpretation — a construction which for most purposes is done spontaneously and without any elaborate apparatus. But in some connections where, for example, we deal with such things as languages, economic systems, or bodies of law, these structures are so complicated that, without the help of an elaborate technique, they can no longer be reconstructed without the danger of going wrong and being led into contradictions.”

13: “The ‘emergence’ of ‘new’ patterns as a result of the increase in the number of elements between which simple relations exist, means that this larger structure as a whole will possess certain general or abstract features which will recur independently of the particular values of the individual data, so long as the general structure (as described, e.g., by an algebraic equation) is preserved. Such ‘wholes,’ defined in terms of certain general properties of their structure, will constitute distinctive objects of explanation for a theory, even though such a theory may be merely a particular way of fitting together statements about the relation between individual elements” ([1964] 2014: 261-2); See also ([1967A] 2014: 281-2) and ([1967B] 2014: 70).


15: In Hayek’s view ([1955A] 2014, 201, fn.3), explanation and prediction are nearly equivalent intellectual tasks.
implicates the sort of invisible hand explanations that Vanberg and others consider to be inimical to Hayek’s group selection account of cultural evolution.

Having explained the meaning of methodological individualism, we can now address the second key idea: the relation between methodological individualism and spontaneous order. In contrast to what Hayek calls “organizations” (1973, p 49), spontaneous orders emerge without commands. Instead, a set of behavior-regulating rules gives rise to an undesigned pattern of interaction. The rules, in this case, are not tailored to achieve some pre-specified goal. Such orders are thus “not the result of all the factors being taken into account by a single centre,” but rather are “produced by the responses of the individual elements to their respective surroundings” ([1955A] 2014, pp 160-1).16 “[I]f a multitude of individual elements obey certain general laws,” Hayek explains, “this may... produce a definite order of the whole mass without the interference of an outside force” ([1955A] 2014, p 160). Whereas the rules that undergird organizations come attached to defined positions with specified goals and serve as complements to commands, the rules that underlie spontaneous orders “must be independent of purpose and be the same... for whole classes of members not individually designated by name” (1973, p 50). Importantly, these rules are “applied by the individuals in the light of their respective knowledge and purposes; and their application will be independent of any common purpose.”17

For Hayek, the abstract nature of these spontaneous orders entails the need for methodological individualism in examining them. Because the interactions between millions of individuals, following only general rules, will exhibit abstract, imperceivable patterns — what Hayek calls “abstract spontaneous orders” (1973, p 38) — we cannot hope to perceive, much less understand, society by examining it only at the aggregate or macro-level. There is no antecedently known pattern which the order strives to realize, and our minds have not evolved to directly perceive abstract, algebraic patterns that are liable to emerge. Although not all spontaneous orders will be abstract (and hence imperceivable), many of them will be, “and for this reason will not be intuitively perceivable and not recognizable except on the basis of a theory accounting for their character” (1973, p 38). In sum, Hayek’s multi-part claim is that (1) many orders within society are spontaneous in the sense that they emerge from rule-following agents, rather than an overarching plan, (2) these orders are often abstract and thus imperceivable, and (3) any adequate account of an abstract spontaneous order will employ a compositive explanation. This type of theorizing — viz. the process of mentally reconstructing a spontaneous order by tracing the relations that exist between the elements — is essentially what Hayek calls methodological individualism.

The economic explanation of social phenomena thus runs from the subjective values, beliefs, or rules that determine individual behavior to the order of actions that emerges when a number of such individuals interact. But, as Wilson and Hodgson are acutely aware, this story contains rather wide gaps. Why should the social scientist assume that the rules followed by individuals will give rise to a spontaneous order, rather than disorder or chaos? And more normatively, given that some orders may be suboptimal or even ghastly, on what basis can we endorse the resulting spontaneous order, as Hayek suggests we should? Vanberg’s criticism of

---

16: Cf. 1973, 43: “the result of their elements following certain rules in their responses to their immediate environment.”

17: This does not imply that there is no common purpose; simply that the individual need not know this purpose in making her decisions. See Mack (2006).
the CASI approach thus seems equally applicable to the CAS2 approach, just as Wilson and Hodgson have claimed. Methodological individualism, like group selection, leaves large explanatory gaps when considered in isolation. As a result, just as group selection fails to satisfy the deductive-nomological criteria of explanation, so does compositive explanation. As shown in the next section (III), to fill these gaps, Hayek turns from the individualistic methodology of pure economics and draws upon multilevel selection theory. The need to account for the explanatory gaps identified in sections II.i and II.ii thus helps to explain why Hayek believed that “nobody can be a great economist who is only an economist” (1956, p 123). By appreciating the integration of these explanatory approaches, Hayek’s social theory can finally be understood.

### 3.3 Reuniting the Twins: Hayek’s Multilevel Selection Theory

As noted above (sec. I), several commentators view Hayek’s dual use of methodological individualism and group selection theory as contradictory; Hayek’s social theory must be modified so as to avoid either CASI or CAS2 explanations. Bruce Caldwell (2004a) presents an alternative view. Although he concurs with the critics that Hayek is inconsistent, Caldwell does not think that Hayek should be viewed as either a CASI theorist or a CAS2 theorist. Rather, Hayek’s enthusiastic endorsement of methodological individualism represents an early phase, and group selection theory eventually superseded Hayek’s early individualism. To support his account of Hayek’s methodological transition, Caldwell cites Hayek’s 1967 paper on cultural evolution, “Notes on the Evolution of Systems of Rules of Conduct,” where Hayek describes how systems of rules will expand or recede to the extent that they affect the “greater or lesser efficiency of the resulting order of the group” ([1967A] 2014, p 279). While Caldwell concedes that individuals still play a role in this process, since it is through their adoption of and adherence to rules that the order obtains its properties, he claims that Hayek’s earlier endorsement of methodological individualism as the only way to understand social phenomena ([1952C] 2010, p 6) was, plainly, a “misstatement” (Caldwell 2004a, p 285).

In Caldwell’s view, it is unimportant whether or not Hayek is a doctrinaire methodological individualist. Caldwell resolves the apparent tension between group selectionist explanations and methodological individualism by simply allowing that sometimes Hayek employs compositive explanations and sometimes group-selectionist ones. Pace Vanberg, Hodgson, and others, Caldwell’s solution does not force a choice. Group selection and the compositive method are simply two alternative tools in the social scientist’s toolbox, perhaps suited to explain different phenomena. There is, however, an issue with Caldwell’s reading: Hayek employs the two tools simultaneously. In fact, the very text that Caldwell cites to establish Hayek’s supposed withdrawal from methodological individualism includes an endorsement of the compositive method:

... it is often easier to ascertain the rules of individual conduct than to trace from them the resulting overall and often highly abstract order...
the ‘functions’ which these rules serve we shall be able to discover only after we have reconstructed the overall order which is produced by actions in accordance with them ([1967A] 2014, pp 281-2).19

Here we see Hayek advocating the compositive method of intellectually reconstructing social orders from an understanding of the actions of individual agents. If group selection and methodological individualism were, indeed, two alternative, mutually exclusive types of explanation, it would be odd for Hayek to explicitly endorse both in the same article. It would be even stranger for Hayek to claim that both are necessary for understanding the nature of social rules. It seems, then, that Hayek does not view these methods as exclusive alternatives. He must hold that they somehow work together.20

As we have seen (II), neither compositive explanations, nor group selection theory taken alone can provide a complete explanation of social phenomena. Moreover, since each of these explanatory methods contributes different elements to the full explanation of social phenomena, they are mutually compatible. Being neither (full) explanations of the same phenomenon, nor mutually exclusive, they do not constitute alternative methods of explanation. Since, taken alone, each type provides an incomplete explanation of the phenomena, Hayek’s claim that methodological individualism is necessary for an adequate understanding of social phenomena is compatible with the claim that group selection is also necessary. Importantly, Hayek does not, and should not, claim that either approach is sufficient to achieve a complete explanation. For Hayek, the two approaches are not, as Caldwell supposes, substitutes; they are complements.

To see how these two explanations work together, recall the major gaps that emerge when each is considered in isolation. Group selection alone leaves us in the dark about how orderly structures emerge from the myopic, rule-following behavior of individuals. Methodological individualism, on the other hand, fails to explain why individuals end up following rules that are conducive to functional orders.

In discussing the concept of spontaneous order, Hayek himself notes that methodological individualism alone fails to explain the “invisible hand” ([1967B] 2014, p 298). Although Hayek’s normative argument does not rely on the claim that currently existing orders are optimal ([1960] 2011, p 48; 1988, p 27; Whitman 1998), it does depend on the claim that the spontaneous orders will generally be beneficial, even if suboptimal. As Wilson and others have pointed out, explanations of social phenomena that employ methodological individualism alone provide no reason whatsoever to believe this (Wilson 2015, pp 95-6). Considering the deductive-nomological theory of explanation, then, the CAS2 approach utterly fails to provide an adequate explanation of functional group order.

Hayek’s solution comes in the form of his theory of group selection. In the context of discussing the invisible hand, Hayek writes,

\[\ldots\text{neither Smith nor any other reputable author I know has ever maintained that there existed some original harmony of interests irrespective of... grown institutions. What they did maintain... was that institutions had developed by a process}\]
of the elimination of the less effective which did bring about a reconciliation of the divergent interests. ([1967B] 2014, p 98).

We expect our grown institutions, the rules that regulate our behavior, to produce beneficial results only because they are the result of a process of group selection. By positing selection at the level of groups — the process detailed in section II.i — Hayek thus provides an account of the functionality (though not the optimality) of the systems of rules that have evolved. We expect these systems to give rise to order rather than disorder simply because a society lacking order will not inspire imitation, will not support wealth creation or population growth, and will be vulnerable to invasion. Hayek thus uses group selection to supplement his individualistic, compositive explanations of social order, explanations which fail to provide a full account without this addition.

Just as methodological individualism requires group selection, so the converse holds. On its own, group selection can explain why the patterns of interaction that arise within groups will be conducive to the functionality of these groups. But it cannot explain what Hayek considered to be the most important object of the social sciences: the nature and existence of spontaneous orders. The special properties of the various spontaneous orders that make up society simply cannot be explained (or perhaps even perceived, as noted above [II.ii]) without additional theorizing. A massive black box mediates between collective functionality and, for example, the properties of the catallaxy or of a particular social morality. Group selection, therefore, fails to provide a true explanation of social phenomena, at least according to the theory of explanation with which Hayek was likely working.

To replace the black box with a bona fide explanation, Hayek invokes compositive explanations. Methodological individualism bridges the gap between rules of conduct and social order — that is, between replicators and vehicles. Evolutionary biology provides an apt analogy here: just as the theory of natural selection explains how certain packages of genes spread throughout a population, so cultural group selection explains how certain systems of rules spread throughout human societies. And just as microbiology explains how variations at the genetic level produce specific phenotypic expressions, so the compositive approach of social science explains how variations at the level of rules give rise to orders of a certain kind with certain properties. To claim that group selection and methodological individualism are mutually exclusive alternatives is as mistaken as claiming that natural selection and microbiology are mutually exclusive explanations of the same phenomena.

The correct account of Hayek’s methodology, therefore, does not eschew either group selection, as Vanberg urges, or methodological individualism, as Hodgson proposes, nor does it treat them as two alternative methods of analysis to be employed separately, as Caldwell suggests. For Hayek, a true explanation of social phenomena requires the full use of both approaches in a mutually complementary fashion. Drawing together these several strands, Hayek’s methodology involves the following logical (not chronological) order of explanation:

1. Multiple competing societies abide by diverse sets of moral/legal rules.
2. Within each society, these rules spontaneously give rise to diverse orders, exhibiting different properties, or they fail to do so.
3. These different properties support different rates of replication; some lead to social dissolution. This occurs in various ways: violent conquest, rates of population growth, or, most importantly, intentional or unintentional imitation of more successful cultures.
4. Through this process of selection, certain systems of rules spread and overtake alternative systems. Such systems of rules could also lead to mutual influence.
5. Variation enters the system: mutual influence between two cultures will lead to ‘breeding’, rules will be unsuccessfully copied by others or imperfectly transmitted to new generations, or innovators will intentionally violate the rules.
6. A new pool of rule-systems thus emerges, and the process is reiterated \textit{ad infinitum}.

Methodological individualism is Hayek’s favored method of bridging steps (I)-(3), while group selection completes the explanation by bridging steps (3)-(6). For a visual depiction of Hayekian explanation, see Figure 1 below.

![Figure 3.1: Multilevel, bidirectional causation](image)

Starting with some systems $S$ of rules, perhaps “stumbled upon” as Hayek puts it (1978, p 164), we have a network of relationships. From this network an order $O$ emerges (or fails to). In the context of alternative systems that also give rise to orders (or non-orders), $O$ bears ramifications for the ability of $S$ to proliferate: a society whose rules generate a more successful order $O$ will, ceteris paribus, successfully compete, $C$, with other societies. That is, their rules will propagate at a quicker rate. This results in a system that, through negative and positive feedbacks $FB$, generates a new population of systems to compete in the next round. Compositive explanations bridge the upper boxes, showing how $S$ leads to $O$ and what properties, beneficial or harmful, the elements of $O$ will exhibit. Cultural evolution explains the process of feedback that occurs as the process “resets” between $C$ and $S$ (at the next period).

Crucially, this feedback system involves mutual causation between the levels. Upward causation (II.ii), studied by applying methodological individualism, occurs as orders spontaneously emerge from agents following rules. Downward causation (II.i), on the other hand, occurs as group-level selection pressures shape the rules that agents in the population adhere to. It is in light of this multilevel, bidirectional causation that Hayek asserts the interdependence of mind and society. Our culture shapes our values and constraints; “reason and civilization develop in constant
Hayek’s Twin Ideas: Reconciling Methodological Individualism and Group Selection

By integrating both upward and downward forms of causation, we arrive at an understanding of Hayek’s social theory that incorporates a bidirectional relationship between individual agency and social structure. Understood in this way, Hayek presents what others have called a transformational model of social order (C. Lawson 1994; T. Lawson 2000). The significance of this result lies in the fact that several other scholars have drawn the same conclusion about Hayek’s social theory, despite approaching his work from completely different angles. Steve Fleetwood (1995) and Jochen Runde (2001), for example, both undertake an analysis of Hayek’s social theory in terms of Tony Lawson’s distinction between empirical realism and critical realism. In doing so, they both conclude that at the heart of Hayek’s theory lies a bidirectional causal relationship between individual agents and social structures. In the context of an entirely distinct inquiry — an inquiry into the significance of emergence in Hayek’s work — Paul Lewis (2015) arrives at a strikingly similar conception. Hayek’s social theory, Lewis concludes, involves “the interplay between. . . two ontologically distinct. . . causal powers, namely, the social order of actions and the power of people to engage in purposeful, creative decision making” (2016, p 1187; see also p 1171). Despite the diversity of our interpretive approaches, Fleetwood, Runde, Lewis, and I converge to the same basic conclusion: far from being mutually incompatible alternatives, group selection and methodological individualism are complementary, even interdependent, parts of a unified, Hayekian approach to understanding social phenomena.

3.4 Conclusion

For two explanations to be alternatives they must fully account for the same phenomenon. For a choice between two alternative explanations to be necessary they must be mutually exclusive. Explanations of social

---

21: Technically, in a rational actor model, enforcement can stabilize these norms only if we have either infinitely repeated interactions, or if we dispense with sub-game perfect equilibrium, relying on mere Nash equilibrium as Binmore (2005: 68-71, 85) does. Andreozzi (2005, pp 237-9) avoids this issue by using replicator dynamics to defend the idea of a cooperative equilibrium with punishers. This relies on assuming simple decision heuristics that defy common knowledge of rationality.
phenomena based on methodological individualism and those that employ group selectionist accounts are neither sufficient on their own, nor mutually exclusive. Therefore, the social scientist need not, and Hayek certainly did not, choose between group selection and the invisible hand accounts of the compositive method. Group selection and methodological individualism are fully consistent, and as a result the "twin ideas" of evolution and spontaneous order can be seen as two aspects of a unified explanatory framework for social phenomenon.

More contestable is whether or not these two levels of adaptation can maintain the functionality of a modern society. One group of theorists believe that CAS2 adaptations suffice to support, at least, an acceptable social order. Others claim that CAS2 adaptations will lead to group dysfunction without efforts to design institutions so that CAS2 adaptations translate into CAS1 adaptations. Hayek's view may be somewhere in-between. Gene-culture coevolution has endowed human beings with rational capacities and has made them rule-following beings by constitution. Yet, humans' deep genetic past also endows them with atavistic urges that threaten to undermine the foundations of modern society unless cultural evolution can succeed in suppressing these primordial instincts. Thus, unlike the CAS2 theorists, Hayek expresses unease at the thought of weak CAS1 pressures. But unlike many of the CAS1 theorists, Hayek relies on a process of cultural evolution, rather than institutional design, to protect collective functionality. Whether or not the pressures that Hayek invokes for this purpose suffice is a question that merits further investigation.
Polycentric Governance
4.1 The Puzzle of Polycentric Functionality

Recent work on polycentricity has clarified its key features (Stephan, Marshall, and McGinnis 2019; Aligica and Tarko 2012). Put succinctly, a polycentric governance structure is one that involves multiple decision-making units, each with authority over a specified, but evolving jurisdiction, which interact in various ways, according to a set of overarching rules. This structure typically engenders diverse forms of competition; for instance, jurisdictions, which may be territorial or non-territorial, often compete for members.

Like hierarchical governance structures, polycentric governance implies that governance occurs at multiple levels and scales. However, unlike hierarchical governance structures, polycentricity is decentralized; it grants significant autonomy to lower-level decision units, and, moreover, these units exert influence on higher level units, producing coordination in a bottom-up fashion (Tarko forthcoming). Decentralization is a feature that polycentricity shares with a fragmented or anarchic institutional structure, but unlike mere fragmentation, decision-making units in a polycentric structure operate under a shared set of overarching rules, allowing them to coordinate (Tarko forthcoming).

Together, these features produce a puzzle regarding the high degree of functionality observed by experts in many polycentric governance structures. Somehow, processes involving both horizontal competition between governance units on the same scale, as well as vertical competition between governance units on multiple scales, generate beneficial social outcomes. Although invisible hand processes often produce a surprising degree of order, they typically presuppose a certain institutional background. Decentralization and competition do not, in general, guarantee functional social order (Wilson 2016). Without the right institutions the invisible hand is liable to become an “invisible fist” (Anomaly and Brennan 2014). Only under the right institutional structure do “invisible hand processes” incentivize local actions that produce global benefits. Why is it, then, that successful polycentric systems succeed? What form must the overarching system of rules take in order for decentralized, interjurisdictional competition to give rise to beneficial order, rather than repugnant equilibria or disorderly chaos?

This paper seeks to advance our understanding of polycentric governance by addressing this question. Our best theoretical tool for understanding the relationship between lower level rationality and emergent functionality is multilevel selection theory. Accordingly, the paper begins by laying out the formal framework of multilevel selection theory, encapsulated in two formulas derived from Price’s equation. Through the lens of this formal framework, the next section shows how the various features of polycentric governance combine to support collective functionality by referencing the possibility of “local tyrannies” that might arise in polycentric, self-governing institutional structures (Ostrom 2009, 282-3).
satisfying these two formulas. Section 4 examines two empirical cases, making it apparent that polycentric political structures alone do not suffice to generate group beneficial outcomes. As Elinor Ostrom has shown, successful governance structures require mechanisms for minimizing opportunistic behavior by would-be rule violators. The final section explores Elinor Ostrom’s eight “design principles” in order to identify the features required to supplement polycentric governance structures, that is, to direct decentralized decision makers to produce global benefits.

4.2 The Price Equation and Multilevel Selection

The Basic Price Equation

When systems evolve as a result of the myopic choices of their members, this generally results “in a breakdown of group-level functional organization” (Wilson, 38). As economists and biologists have long known, lower-level rationality does not automatically scale. Social dilemmas abound, leading to suboptimal outcomes for economies and species. In various ways, social planners might avoid such outcomes by directing the behavior of individuals towards socially desirable ends, but this solution can only work in extremely simple societies. In more realistic settings, institutional frameworks allow for decentralized decision-making, polycentric governance structures offering an extreme example of such frameworks. Given the high degree of local autonomy, we must wonder how polycentric governance structures are able to secure desirable outcomes at the global level. Individuals within various jurisdictions do not, in general, possess the knowledge or the desire to produce group-beneficial outcomes. Yet, somehow, polycentric political organization often promotes effective governance (Thiel, Blomquist, and Garrick 2019). What features of polycentric organization underlie this surprising capacity?

To understand how a decentralized network of decision-making entities can produce and maintain beneficial social order, it will help to draw on multilevel selection theory. In particular, two versions of the Price equation provide deep insight into the conditions in which such social order will arise and persist.

First, we need to establish some notation:

- $z$: A (phenotypic) trait of some kind, which can be measured (discretely or continuously) with real numbers.
- $z_i$: The level of $z$ exhibited by the entity $i \in P = \{1, \ldots, n\}$.
- $\bar{z}$: The average level of $z$ in the $P$-population.
  - Mathematically, $\bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$
- $\Delta z_i$: The change in the level of $z$ from one entity to its offspring.
  - Taking an average across all entities in $P$, the expectation of $\Delta z_i$ is $E[\Delta z] = \frac{1}{n} \sum_{i=1}^{n} \Delta z_i$
- $w_i$: The “fitness” or average number of offspring produced by entity $i$. 
• Similarly to $\bar{z}$, we define the average fitness as $\bar{w} = \frac{1}{n} \sum_{i=1}^{n} w_i$.
• Often we will be interested, not in the absolute fitness of an entity, but in its fitness relative to the other entities in the group. For this comparative purpose, we define relative fitness of entity $i$ as $\omega_i = \frac{w_i}{\bar{w}}$.

With this notation in hand, a form of the Price equation can be written as follows:  

$$\Delta \bar{z} = Cov(\omega, z) + E[\omega \Delta z].$$

(4.1)

Although simple, (Price Eq.) reveals the key components of evolutionary change. It decomposes the total change in $\bar{z}$, that is, the total change in the average level of the $z$-trait, into two components: $Cov(\omega, z)$ and $E[\omega \Delta z]$. $Cov(\omega, z)$ denotes the statistical association between $\omega$, relative fitness, and $z$, the level of the $z$-trait. When these two variables move together, that is when greater relative fitness is associated with higher levels of $z$, the covariance will be positive: $Cov(\omega, z) > 0$. When they are totally unrelated, $Cov(\omega, z) = 0$. And when they move in opposite directions, e.g. higher levels of $z$-trait go along with lower relative fitness, the covariance will be negative: $Cov(\omega, z) < 0$. Therefore, this first term is often identified with the evolutionary force of selection (Gardner 2008).

The second quantity on the right-hand side of equation (Price Eq.), $E[\omega \Delta z]$, denotes the weighted average of the change in $z$-levels between parents and offspring. Each entity $i \in P$ exhibits level $z_i$ of the $z$-trait and produces some number of offspring $w_i$ with a level $z_i'$ of the $z$-trait. If we want to measure the average transmission rate of the $z$-trait, it makes sense to take the average change between parents and offspring, but for a more accurate measure, we should also heavily weight those entities with many offspring and discount those with few offspring. $E[\omega \Delta z]$ achieves this weighting by multiplying each $\Delta z_i$ by the relative fitness of entity $i$, which is $w_i/\bar{w} = \omega_i$. In short, $E[\omega \Delta z]$ gives us a measure of the population’s overall transmission bias or copying fidelity. If all entities in $P$ produce offspring with higher levels of $z$-trait, then $E[\omega \Delta z] > 0$. If they all tend to produce entities with lower levels of $z$-trait, then $E[\omega \Delta z] < 0$. When the transmission is perfect, i.e. $z_i = z_i'$ for all $i \in P$, then $E[\omega \Delta z] = 0$. In that case, the only evolutionary force in operation is selection.

To achieve some intuition for these terms, let us consider a specific interpretation. In this interpretation, we consider two periods, one where jellyfish compete for resources, and another in which these jellyfish reproduce. We also assume that the $z$-trait exerts direct causal force in determining the survival of these jellyfish, but not in determining their fecundity. In this simple scenario, $Cov(\omega, z)$ tells us how well the $z$-trait promotes survival in the first period. Some jellyfish will die, some will survive, and $Cov(\omega, z)$ tells us how much the $z$-trait has contributed to survival ability. In the second period, when the jellyfish reproduce, $E[\omega \Delta z]$ adds a further change by representing the amount of the $z$-trait inherited by the offspring.

This decomposition seems rather straightforward: some part of evolution will be due to the relation between fitness and the $z$-trait, and some part of it will be due to the ability of entities to actually bestow the
z-trait to their offspring. We might think of the first term, \( \text{Cov}(\omega, z) \) as representing the basic evolutionary idea that if a trait helps an entity survive and reproduce, it will proliferate. The second term, \( \mathbb{E}[\omega \Delta z] \), adds the obvious qualification that a beneficial trait will spread only if parents actually pass it on to their offspring.

The standard way of writing the Price equation takes equation Price Eq. and multiplies through by \( \bar{w} \) to yield:

\[ \bar{w} \Delta \bar{z} = \text{Cov}(w, z) + \mathbb{E}[w \Delta z] \]  

(4.2)

Notice that the only real change is that \( \omega \) has been replaced with \( w \). This is just because \( \omega_i = \bar{w}_i / \bar{w} \), and we multiplied through by \( \bar{w} \). From this basic form of the Price equation, we can derive a simple expression that models multilevel selection.

### The Cultural, Multilevel Price Equation

The Price equation purports to be an entirely general description of any evolutionary process. It should, therefore, extend to describe a process of cultural, multilevel selection, e.g. institutional evolution within a polycentric political framework. To see this, we need merely offer an interpretation of equation A.2. Suppose we have a set of groups indexed by \( j \in \{1, \ldots, N\} \). For any individual \( i \) in group \( j \), the trait \( z_i \in \{0, 1\} \) represent adherence to a rule or institutional feature that is “altruistic” in a technical sense. That is, \( z_i \) represents adherence (1) or non-adherence (0) to a rule that promotes the well-being of the group \( j \) (defined as the average individual well-being within group \( j \)), but which decreases that of the individual \( i \). Accordingly, \( z_j \in (0, 1) \) will represent the average level of the altruistic trait within group \( j \), and the variable \( w_j \) indicates the average fitness within group \( j \). That is, how many copies of the rule will members of group \( j \) produce on average? Though having more offspring may lead to more copies of a rule, increased copying of a rule does not entail increased offspring or population increase. Instead, \( w_j \) measures the number of other individuals that will copy group \( j \) by imitating group \( j \)’s rule level \( z_j \).

To take an example, \( z_i \) could represent a norm of respecting the personal property of others. As an individual, \( i \) could enhance her material condition by violating this rule, yet such a rule benefits the average member of \( i \)’s group. Consider two groups, \( j \) and \( k \). Each group has 10 members. Group \( j \) mostly adheres to property rules, only two out of its ten members are thieves. Group \( k \) rejects property rules, only two out of its ten members respect property. To derive results form this, suppose that each individual follows a policy of “prestige bias” or copying the successful. And each individual \( i \) has the following utility function, characteristic of an altruistic trait:

\[ u_i = \alpha z_j - \beta z_i \]  

(4.3)

Will the property rule spread or recede? Given the initial state of the population divided into groups \( j \) and \( k \), the average payoff for those...
following the property rule will be the weighted sum of the average in group $j$ and that in group $k$:

$$0.8[a(8) - c] + 0.2[a(2) - c] = 0.68a - c$$

Meanwhile, the average payoff of rule violators will be:

$$0.2[a(8)] + 0.8[a(2)] = 0.32a$$

Thus, in this initial state, the property rule will be drawing more adherents if and only if . . .

$$0.68a - c > 0.32a$$

$$0.36 > rac{c}{a}$$

In other words, if the benefits of living in a group of rule followers significantly outweighs the personal cost of adhering to a property rule, then the rule will tend to spread.\(^7\)

The basic reasoning governing this simple example can be represented analytically by another version of the Price equation:\(^8\)

\[\bar{w}\Delta \bar{z} = \text{Cov}(z_j, w_j) + E\left[\bar{w}_j\Delta z_j\right]\] \hspace{1cm} (4.4)

\[\bar{w}\Delta \bar{z} = \text{Cov}(z_jw_j) + E\left[Cov(w_i\Delta z_i) + E\left[w_i\Delta z_i\right]\right]\] \hspace{1cm} (MLPE)

The addition of the subscript $j$ in equation (4.4) indicates that we are considering group quantities: $z_j$ is the average level of trait $z$ within the group and $w_j$ is the average fitness of the group. The addition of the subscript $i$ in the next line, (MLPE) indicates that we are considering within-group covariances and expectations, taken over individuals within a fixed group $j$. As above, the term $E\left[w_i\Delta z_i\right]$ expresses non-selective forces, such as drift, transmission bias, or random effects. To focus on evolutionary forces we therefore ignore this term by setting it equal to 0.

The next subsection reveals some important implications of this version of the Price equation. In particular, two equations derived from (MLPE) provide a clean representation of the conditions under which group beneficial rules are likely to evolve.

**Implications of the Price Equation**

Two derived formulas illuminate the implications of the Price equation for evolved group functionality. To derive these, first note that, because $z_j$ and $w_j$ are themselves averages, the second term in equation (MLPE) can be decomposed in the same exact way as we decomposed $\bar{w}\Delta \bar{z}$. Doing so yields:

One more mathematical fact is required to complete the derivation. If we denote as $\beta_1$ the regression coefficient for $w_j$ on $z_j$, then we have and $\beta_2$ the regression coefficient for $w_{ij}$ on $z_{ij}$:\(^9\)

\[\text{See appendix for a more thorough explanation.}\]
Polycentricity and Adaptation
A Multilevel Selectionist Approach

This second formula is inspired by Henrich (2004), but see the appendix A.

Letting $E[w_{ij} \Delta z_{ij}] = 0$ for reasons discussed above and substituting these two equations into (MLPE) produces the following:

\[ \tilde{w} \Delta \tilde{z} = \beta_1 V(z_i) + \beta_2 V(z_{ij}) \]  \hspace{1cm} (4.5)

Because they are regression coefficients, $\beta_1$ gives a measure of how changing the average level of $z$ within a group affects the average fitness of the group, while $\beta_2$ measures how changing the level of $z$ of an individual within a group will change that individual’s fitness. The expression (4.5) thus represents the evolution of the average level of trait $z$ as a composition of a between group portion and a within-group portion. From (4.5), we can infer that the trait $z$ will spread when...

\[ \frac{V(z_i)}{V(z_{ij})} > -\frac{\beta_2}{\beta_1} \]  \hspace{1cm} (*)

Assuming that the trait $z$ is “altruistic” in the technical sense of benefiting the group, but not an individual within a group, then $\beta_1 > 0$ and $\beta_2 < 0$.

There are several crucial implications of the expression (*). The left-hand side states that a prosocial trait $z$ is more likely to spread when variability within groups is minimized and between-group variance is maximized. The right-hand side tells us that the strength of selection pressures is also crucial. If the trait $z$ is extremely harmful to an individual within a group, i.e. $|\beta_2|$ is large, then it is unlikely to evolve. If it is extremely beneficial for the group, i.e. $\beta_1$ is large, then it is more likely to evolve.

This framework allows for a rigorous analysis of the sorts of design features that will enable group-beneficial adaptations, even when such adaptations run against lower-level selective pressures.

The Price equation also suggests another useful formula that can aid in understanding what conditions must be met for a group beneficial rule to proliferate. As a first step, consider breaking up and individual’s fitness, $w_i$, into two components, one determined by the rule itself and the second determined by the amount of rule adherence within the individual’s group or, more precisely, network of interaction. Each of these components will make some separate contribution, but they are not statistically independent. Having a high level of trait $z$ may, for example, predict that one’s network of interaction is more likely to exhibit high average levels of $z$. This is not only because an individual will directly contribute to the average level of $z$ within his or her group, but also because individuals with high levels of $z$ may preferentially interact with others who exhibit a high (or low) level of $z$. To isolate the effects of individual $z$ levels from those of group $z$ levels, we must therefore write out $w_i$ as a sum of partial regression coefficients (Allen 1997). To denote the fitness effect of increasing an individual’s level of rule adherence, $z_i$, while holding group adherence constant, we write the partial regression coefficient $\tilde{\beta}_{w_i z_i}$. Similarly, to denote the fitness effect...
of increasing group adherence, $x_j$, while holding individual adherence constant, we write the partial regression coefficient $\beta_{w_i z_j}$. Putting these together in a regression equation yields...

$$w_i = \beta_0 + \beta_{w_i z_j} x_i + \beta_{w_i z_j} z_j + \epsilon$$

where $\beta_0$ represents base fitness and $\epsilon$ is an error term. Substituting (A.6) into the Price equation produces the following equation:\footnote{See the appendix A.3 for a thorough explanation and derivation of this equation.}

$$\bar{w} \Delta \bar{z} = \beta_{w_i z_j} V(z_i) + \beta_{w_i z_j} \beta_{z_j z_i} V(z_i)$$

$$= (\beta_{w_i z_j} + \beta_{z_j z_i} \beta_{w_i z_j}) V(z_i)$$

Since $V(z_i) \geq 0$ as a mathematical fact, this implies that a trait will be selected for, i.e. $\bar{w} \Delta \bar{z} \geq 0$, only if we have:

$$\beta_{w_i z_j} + \beta_{z_j z_i} \beta_{w_i z_j} > 0 \quad (**)$$

If we assume we are talking about an altruistic trait, then we know the following facts:

$$\beta_{w_i z_j} < 0$$
$$\beta_{z_j z_i} > 0$$

Given these two facts, our prediction of whether an altruistic trait will spread depends upon a crucial structural feature of the population-interaction structure. In order to ensure that condition (**) is satisfied, we would like $\beta_{z_j z_i}$ to be large and positive. In other words, going back to the intuitive meaning of the expression, an altruistic trait is more likely to be selected when altruists are capable of bunching together. This point is of fundamental importance for understanding multilevel selection and the evolution of altruism, so it bears repeating: in order for an altruistic trait to evolve (through selection), altruists must have some mechanism(s) for excluding egoists from their network of interaction or, equivalently, of converting egoists within their network into altruists.

Summing up the implications of (*) and (**), we can identify four functionality desiderata:\footnote{Because these terms are important only relative to one another, these four desiderata could be reduced to two desiderata – or even to a single desideratum if we wanted to be fully parsimonious. I separate them here for analytic clarity, but we must remember that each of the desiderata must be appended with a certis paribus clause. I thank Vlad Tarko for pointing this out.}

1. Prosocial traits are more likely to emerge when within-group variance $V(z_i)$ is minimized, perhaps due to the ability of altruists to group together and to exclude or convert egoists.
2. Prosocial traits are more likely to emerge when within group selective pressures against the trait ($|\beta_{z_j z}|$) are small. Prosocial traits that do not require extreme sacrifice are thus more likely to proliferate, but this might also involve setting up institutional mechanisms to punish egoists and to reward altruists.
3. Prosocial traits are more likely to emerge when between group selective pressures for altruism ($\beta_{1}$) are large. In times of frequent and intense interaction, especially, some believe, when resources are scarce and interactions are agonistic, high levels of cooperation at the social level are imperative.
4. Finally, prosocial traits are more likely to emerge when variance between groups $V(z_i)$ is large.
These four conditions provide the key for understanding the spectacular successes – and failures – of polycentric governance arrangements. In the next section, we will demonstrate how the features of polycentric political organization help to fulfill several of these conditions. But we must also compare cases in which polycentricity succeeds to cases in which it fails, so as to provide an account (sec. 4.5) of variable levels of success in polycentric governance structures.

**Group Fitness and Group Welfare**

Before applying this formal framework to the analysis of polycentric governance structures, a conceptual issue requires clarification. The Price equation tells us what will make a rule adaptive at the global level. Rules that raise the average fitness of individuals within a group will spread to individuals in other group via imitation, immigration, conflict, or some combination of forces. The question at issue, however, is why polycentric competition often provides good governance, i.e., why it often promotes social welfare. But is there any reason to suppose that group fitness corresponds to group welfare?

Many have argued forcefully against this supposition. James Buchanan, for instance, has accused F.A. Hayek of adhering to the Panglossian fantasy that whatever evolves must be desirable. “My basic criticism of F.A. Hayek’s profound interpretation of modern history and his diagnoses for improvement is directed at his apparent belief or faith that social evolution will, in fact, insure the survival of efficient institutional forms” (Buchanan 1975, 211). Buchanan holds that Hayek’s position amounts to normative evolutionism, that is, that we should passively accept the outcome of any cultural evolutionary process (Buchanan 1986, 312). In a similar vein, Dan Dennett has criticized a host of normative evolutionists for failing to explain why evolutionary outcomes should correspond to consciously chosen human values (Dennett 1995, 468). Would survival of one’s culture “justify mass murder, for instance, or betraying all your friends?” Dennett asks. Clearly not. So, why, then, should we associate evolutionary success with value or desirability?

To understand why cultural evolution within a polycentric system avoid such criticisms requires a careful analysis of the meaning and operation of polycentric orders, as well as additional features that such orders require in order to produce group beneficial outcomes. Undertaking this analysis occupies the remainder of this paper. However, a succinct response draws on the fact that cultural selection, as understood here, posits that selection occurs in such a way as to raise the average, relative fitness of a certain system of rules. Clearly, there are certain ways in which this striving for greater relative fitness can come at the expense of human welfare, e.g. through violent conflict. Just as obviously, however, greater relative fitness can be pursued in ways that enhance group welfare, e.g. through cultural or economic competition. Polycentric orders are, crucially, governed by an overarching set of rules (Tarko forthcoming). Such rules will, in general, aim to reduce non-productive, zero-sum competition, while, at the same time, promote rivalry that leads to useful institutional experimentation. In this way, the overarching set of rules sets up a framework for evolution that supports a correlation between evolutionary fitness and human welfare. Absent an overarching set of

---

13: One response, put forth by Hayek on various occasions, is that our deepest values are themselves products of cultural and biological evolution. For instance, Hayek writes that “value ... can only be understood as the determinant of what people must do to maintain the overall structure” (Hayek 1983, 36). This response raises a host of difficulties and complexities, which, if left unaddressed, render it unconvincing. Although I believe there is some merit to this response, laying it out in sufficient detail would lead us far afield.
rules, the system is not polycentric, but fragmented and anarchic (Tarko 2016, 43). There is also the possibility that the overarching set of rules will be dysfunctional. As we will see, polycentricity alone does not suffice to ensure beneficial outcomes. Nevertheless, it does possess several desirable properties, which, when supplemented by certain design features, greatly increase the likelihood of beneficial evolutionary outcomes.

4.3 Polycentricity and Multilevel Selection

Defining Polycentricity

Vincent Ostrom offers a concise and now classic definition of polycentricity:

A polycentric political system [is] composed of: (1) many autonomous units formally independent of one another, (2) choosing to act in ways that take account of others, (3) through processes of cooperation, competition, conflict, and conflict resolution. (V. Ostrom 1991, 225)

Other scholars have built upon this definition to provide greater precision. Especially notable are two teams of scholars. First, Aligica and Tarko (2012, 257) present a “concept design” that involves three key features of polycentricity, as well as a host of empirical indicators for each of these features. The three features are:

1. A multiplicity of decision centers
2. An overarching system of rules
3. A process of evolutionary competition between the decision centers.

A second team of scholars—Stephan, Marshall, and McGinnis (2019)—provide a list of eight features of polycentric systems. However, they consider four of these features to be of special importance:

1. Multiple decision centers
2. Autonomous decision-making authority for each decision center
3. Overlapping jurisdictions of authority between the decision centers
4. Various processes of mutual adjustment among decision centers

While several others have also provided definitions of polycentricity, they all more or less resemble the three definitions covered here.

Pulling together the various features of these three different definitions, we can identify a basic schema for the organization of polycentric governance:

**Polycentric Political Structure**: A polycentric political structure consists of rule-governed collectives with well-defined, and often overlapping jurisdictions that interact in a rule-governed, competitive manner resulting in the relative expansion or contraction of their jurisdictions.

To reduce this definition to an orderly list, we might identify three features:

1. Multiple decision-making units
2. Each decision-making unit has authority over a specified, but evolving jurisdiction.
3. These decision-making units interact in various ways, according to a set of overarching rules, as they compete for members.

This precise understanding of polycentricity enables an assessment of polycentric functionality in terms of the conditions for group functionality laid out above, an assessment to which we now turn.

**Facilitating Multilevel Selection**

The key to understanding how polycentric political organization enables beneficial rivalry is to juxtapose the above definition of polycentricity with the conditions for prosociality derived above. Recall, in particular, the second two functionality desiderata that follow from expressions (\(\ast\)) and (\(\ast\ast\)):

3. Prosocial rules are more likely to emerge when between group selective pressures (\(\beta_1\)) are large.
4. Prosocial rules are more likely to emerge when variance between groups \(V(z_j)\) is large.

The first and second features in our definition of a polycentric political structure favor prosociality by increasing \(V(z_j)\). Boundaries between groups are crucial for developing distinctive sets of rules. Institutional diversity presupposes distinct jurisdictions. Moreover, without well-defined groups, all adhering to the same rules, the variance between collectives will either be less pronounced, or undefined, since the collectives themselves will be undefined. Between groups, in other words, looser boundaries make for lesser variance.

The third feature in our definition of a polycentric political structure favors prosociality by increasing \(\beta_1\). When jurisdictions with diverse rule sets are in a state of constant interaction, members of other jurisdictions become familiar with the alternative rules and with their effects with respect to well-being. If the jurisdiction is geographical, they are more likely to engage in *Tiebout competition* by “voting with their feet.” If the jurisdiction is non-geographical, individuals may simply switch to the governance provider who yields better results at lower costs. In this way, interaction between jurisdictions intensifies group-level selective pressures. Rules that fail to enhance human well-being (or, at least, seem to fail) thus face more rapid decline than they would under autarkic conditions with lower levels of interaction.

What about the within-group components, i.e., the first two functionality desiderata?

1. Prosocial traits are more likely to emerge when within-group variance \(V(z_{ij})\) is minimized.
2. Prosocial traits are more likely to emerge when within group selective pressures against the trait (\(|\beta_2|\)) are small.

These elements are not addressed by polycentric organization as directly as the second two are, an issue that plays an important role in sections 4.4 and 4.5. However, by focusing on the second condition (\(\ast\ast\)), it is possible to glimpse at least one way in which polycentric governance...
favors prosociality by affecting within-group components. Recall that condition \((\ast)\) says, in essence, that prosocial traits spread easier when there is some way of excluding egoists from a network of prosocial interactors, or of converting egoists within a network.\(^{14}\)

Polycentric organization facilitates the satisfaction of \((\ast)\) by allowing like-minded individuals to coalesce into groups centered around shared concerns. As Paul Aligica (2018, 104) has argued, polycentric political structures enable the spontaneous formation of groups in which individuals share a sense that they are cooperating in a “problem solving” context. Having identified a pressing problem, citizens are more likely to accept compromises, do their fair share, and censure violators. Importantly, violators are more likely to respond positively to censure, since they accept the basis of social rules and recognize their importance for solving an imminent problem. These features of a problem solving context, facilitated by polycentric organization, provide reason to believe that individuals within the group are more likely to adhere to the social rules on a similar level. Within group variance, in other words, is minimized by the spontaneous formation of problem solving contexts permitted by the polycentric political organization.

Polycentric political organization thus seems to support group success by promoting the satisfaction of functionality desiderata \((1), (3), \) and \((4)\). But what of desideratum \((2)\)? Here, it seems, polycentric organization has little bearing. A governance system might be decentralized, comprised of interacting jurisdictions engaged in solving pressing problems, while still exhibiting rules that fail to minimize within-group fitness differentials.\(^{15}\)

The reason for this is that polycentric governance is a formal feature which does not specify the exact rules selected and enforced within each jurisdiction. Within any given jurisdiction, therefore, the rules in force may provide huge relative benefits to violators and impose large costs on conformists. This is no trivial matter, since it concerns the evolutionary stability of group-beneficial rules (Dawkins 2016; Williams 1996). To overcome complete the account of group functionality, then, something in addition to polycentric governance structures is required. The necessary addition will place constraints on the rules in force within jurisdictions. Before identifying the exact type of constraint necessary, it is worth considering a pair of real-world examples in order to contrast a case of successful polycentric governance with an unsuccessful case.

### 4.4 Examples: Success and Failure

To understand why polycentricity aids in producing group beneficial outcome, but also why it is an insufficient condition for such an outcome, it helps to consider real-world cases. As noted above, polycentricity in itself says little about desideratum \((2)\), that of minimizing within-group fitness differentials. As we shall see, this is a direct consequence of its omission of institutionalized punishment. Without punishment, it is difficult to prevent free-riders from accruing large benefits at the expense of those who comply with the rules.

---

14: This is, of course, quite similar to one of the conditions that helps to satisfy \((\ast)\), i.e. reducing the within-group variance \(V(z_{ij})\).

15: It’s important to bear in mind that jurisdictions may or may not be territorial. Religion, for example, has been analyzed as a polycentric order and its jurisdictions are defined by its members, not by a geographical area (Gill 2020).
CAMPFIRE Project Zimbabwe

An illuminating case of polycentric governance emerged in Zimbabwe in the late 1980s. The Communal Areas Management Program for Indigenous Resources (CAMPFIRE) sought to address wildlife management issues by devolving government functions to lower level, local units. Prior to CAMPFIRE, wildlife was legally the property of Zimbabwe’s federal government, which attempted to pursue conservation objectives by restricting licenses for use (Frost 2008). For various reasons, this scheme produced conflict and dysfunction. A major issue was that villagers, living side by side with the purportedly valuable wildlife, viewed African fauna as pests or even as an imminent danger. A second issue was the opportunity cost, born by relatively poor citizens, who were expected to forgo greater crop yields and supplementary protein for the sake of abstract government objectives. For these reasons, government bans on killing wildlife or expanding cultivation meant that “villagers came to feel that government wildlife protection was persecuting people for the sake of the animals. And they were basically right” (Schmidtz 1997, 334). Compounding these conflicts was the dysfunction of a government struggling to monitor and enforce its policies on an unwilling populace (Matzke and Nabane 1996, 70).

CAMPFIRE radically changed wildlife management in Zimbabwe by transferring wildlife ownership and management to the local level. Nearly 40 out of Zimbabwe’s 57 districts now govern their own wildlife policies. The governance structure is nested and complex (Frost and Bond 2007, 778-9). Each district has an elected Regional District Council, responsible for selling hunting and tourism rights to entrepreneurs. These councils then distribute the profits to citizens in their district or invest them in projects such as schools, clinics, water wells, irrigation pumps, or other community improvements (Frost and Bond 2007, 782). Instead of the federal government regulating hunting according to a single policy, different jurisdictions now set their own policies in a decentralized fashion. The result is that villagers have revised their relationship with wildlife. Having a stake in the direct payments it brings to the community as well as the tourism industry that it fosters, they view it as more than a mere nuisance. Consequently, the monitoring and punishment problem has been largely resolved. As one Zimbabwean official put it, “everyone in the community is a policeman now” (quoted in [Schmidtz 1997]).

Ostrom’s Cases of Failure

In Governing the Commons, Elinor Ostrom identifies five cases of failed community governance and several more cases that she deems “fragile” (Ostrom 1990, 180). Four out of the five cases of failed governance exhibit a similar problem to that confronting Zimbabwean wildlife management prior to CAMPFIRE, namely, an inability to monitor and punish rule violators. One of Ostrom’s rich case studies describes a fishery in the Bay of Izmir on the West coast of Turkey. The various fishing cooperatives located near the bay had no agreed-upon rules to adjudicate their disputes or to limit access to the exhaustible fish population. There did exist “official” rules intending to protect the fishing resources, but “[t]he agency responsible for fishery rules... employed no agents to enforce
those rules” (Ostrom 1990, 146). Nor did the coast guard, police, or any other agency prioritize the enforcement of these rules. As fishers entered the market, productivity declined sharply; the consequence was a textbook case of the *tragedy of the commons*.

A similar story can be told about many of Ostrom’s other examples of failed governance. In some of these cases, the relevant communities fail to formulate and agree upon rules to be enforced. In managing a groundwater basin in San Bernardino, CA, for example, an extreme diversity of conflicting interests thwarted convergence on a set of shared rules (Ostrom 1990, 149). In others cases, however, even widely recognized rules are impossible to enforce. Despite having agreed upon rules which were successfully in force for many years, a fishery in Mawelle, Sri Lanka ultimately failed to enforce quotas on the number of fishing nets used in the area (Ostrom 1990, 155-7).

**The Importance of Monitoring and Enforcement**

Many factors are at play in determining the success or failure of a given governance arrangement, but the ability to enforce rules constitutes a common and decisive thread. Returning to our functionality desiderata, detection and punishment directly affects the value of $\beta_2$. Punishment, after all, aims precisely to ensure that deviant behavior does not pay off. In a system of costless information and punishment, $\beta_2$ would be 0; real-world punishment systems content themselves with driving $\beta_2$ as low as efficiency allows.\(^\text{17}\)

The CAMPFIRE program properly aligned incentives by turning the best-informed agents into the residual claimants of efficient resource management. This enabled successful monitoring and enforcement of rules. By contrast, in Ostrom’s example of the Turkish fishery, those responsible for monitoring and enforcement had little stake in the system. Moreover, compared to the fishers themselves, government agents possessed poor and incomplete information about which fishers were violating which rules. Consequently, rule violations were costly to monitor, and rules were simply not enforced.

The upshot of these case studies, then, is to suggest that a successful governance structure, polycentric or otherwise, must successfully detect and punish rule violators. In the multilevel selection framework, this is necessary in order to minimize $|\beta_2|$. This conclusion was arrived at by Elinor Ostrom herself, but not through the use of a theoretical framework like the multilevel selection framework relied upon here. Instead, Ostrom arrived at this conclusion through a careful analysis of an array of real-world cases. The result of her analysis is a set of eight design principles.

**4.5 Fortifying Polycentric Governance**

Polycentricity, we have seen, supports group-beneficial outcomes, but is on its own insufficient to guarantee them. In particular, polycentric governance structures that lack adequate means of detection and punishment will fail to facilitate a process of selection that prioritizes group
benefits over individualistic opportunism. Elinor Ostrom’s eight design principles express much the same conclusion. She lays them out as follows:

(DP1) Boundaries are clearly defined  
(DP2) Rules conform to local conditions  
(DP3) Agents affected by the rules can participate in modifying them  
(DP4) Monitoring is successful in detecting rule violators  
(DP5) Punishment occurs in a graduated manner, depending on the severity of the violation  
(DP6) There are low-cost arenas for resolving conflicts and disputes  
(DP7) External governmental authorities recognize the rights of communities to devise their own institutions  
(DP8) There are multiple layers of nested enterprises, including private actors and governance institutions

While conditions (DP1), (DP7), and especially (DP8) are constitutive of polycentric structures, the other conditions present additional desiderata that fortify the efficacy of polycentric governance structures. Yet again, they do so by affecting the parameters in (⋆) and (⋆⋆), thus satisfying the functionality desiderata.

Condition (DP2), for instance, will perform the dual function of reducing \(|\beta_2|\) and increasing \(V(z_i)\). By ensuring that local rules are not determined by a “one size fits all” policy, individual sacrifice involved in rule following \((|\beta_2|)\) will be minimized. By enabling diverse conditions to generate diverse sets of rules, \(V(z_i)\) will increase, as different sets of rules emerge throughout the larger system.

Condition (DP3) seeks to motivate rule adherence by granting individuals some power in selecting the rules. It thus reduces \(V(z_{ij})\), since agents are more likely to follow the same rules when they view themselves as having co-authored them. Rules imposed from without lack effective authority, and are less likely to inspire compliance. It may also contribute to minimizing \(|\beta_2|\), insofar as rules that demand extreme sacrifice from certain citizens will likely face extreme opposition in collective choice procedures.

Conditions (DP4) and (DP5) constitute the most crucial supplement to polycentric governance institutions. As we saw in the theoretical discussion (sec. 4.3), polycentric organizations straightforwardly address functionality desiderata (1), (3), and (4), but leave the crucial desideratum (2) undetermined. As our empirical examples indicate (sec. 4.4), desideratum (2), which concerns the individual, within-group cost of rule compliance \((|\beta_2|)\), must be satisfied in order for group functionality to emerge. It is therefore no surprise that Elinor Ostrom observed that successful governance structures must detect and punish rule violators. Failing to do so leads to ineffective governance and dysfunctional societies.

Condition (DP6) will also work to minimize \(|\beta_2|\). When low-cost adjudication is available, rule-following becomes easier for two reasons. First, the relevant rule becomes less ambiguous, as a third party is drawn upon to identify it. Second, the role of a competent judge is, at least, to identify and articulate a rule that makes compliance less costly than conflict. Thus, when a legitimate conflict emerges, the gains from cooperation

18: The analysis in this section complements that found in Wilson, Ostrom, and Cox (2008). The key innovation is to explicitly discuss the importance of the design principles in the framework provided by the Price equation.

19: Subsequent work has extended Ostrom’s principles either through late additions or by separating out certain principles into component sub-principles. See Tarko (2016) and Cox, Arnold, and Tomás (2010) for updated lists of Ostrom’s design principles.

20: This is an empirical claim, one that Ostrom collected ample evidence to support. See Ostrom (2006). The term “effective authority” comes from (Morris 2002, 211).
are unlikely to fall entirely to one party. Of course, this is not to say that when the rule is already recognized and clear-cut, with one party unambiguously in the wrong, that resolution will involve splitting the gains. Instead, it simply means that when rules must be identified, revised, or articulated in ambiguous cases, third parties are unlikely to impose all costs on one party and to confer all benefits on the other. Moreover, since even unfavorable decisions will often be less costly than continued conflict, condition (DP6) will promote rule adherence.

Elinor Ostrom’s design principles thus fortify polycentric governance structures against undesirable evolutionary tendencies. Most importantly, as our brief discussion of (DP2)-(DP6) has revealed, they work together to satisfy functionality desideratum (2), viz. the minimization of $|\beta_2|$.

### 4.6 Conclusion

Drawing on the theory of multilevel selection, this paper has addressed the question of why polycentric political structures exhibit enhanced group functionality. Polycentricity allows a system to satisfy several functionality desiderata that promote the evolutionary emergence of group beneficial outcomes. Not all polycentric structures are highly effective, however, and there must therefore exist an additional set of conditions that fortifies their ability to provide effective governance. From a theoretical perspective, the Price equations suggests that polycentric governance requires an additional component, one that operates so as to reduce the within-group benefit of rule-violating behavior. Empirical case studies confirm this theoretical insight, and several of Elinor Ostrom’s design principles can be understood precisely in terms of the need to reduce the benefits of rule violation. Group functionality can hardly emerge amidst ubiquitous defection and free-riding.
The Problem of Complexity and the Emergence of Polycentric Political Order

5.1 Introduction

A plethora of recent work by political scientists has made impressive progress in clarifying the concept of polycentricity as first expounded by Michael Polanyi (1951) and brought into mainstream political science by Vincent Ostrom (1972). Since the time of Polanyi and Ostrom, a number of important studies have elucidated various aspects of polycentric governance structures (Aligica and Tarko 2012, 2013; Aligica 2014, 2018; Thiel, Blomquist, and Garrick 2019; Vogler 2020; Müller 2019; Kogelmann 2017; Gaus 2016, 2021). As a result of this large and growing literature, social scientists and political theorists now have a deeper understanding of the properties, benefits, and limitations of polycentric systems. Despite its richness, this discussion exhibits a conspicuous gap. Political scientists have yet to closely examine the emergence of polycentric political systems. This is unfortunate for at least two reasons. First, the historical emergence of polycentric political systems is of intrinsic interest to those who study economic or political history. And second, understanding how and why polycentric political systems emerge can shed light on theoretical questions by providing additional lines of support for preceding analyses or suggesting new hypotheses about the nature of polycentrism. The present article aims to fill this gap by proposing an explanation of polycentric state emergence.

This article begins by carefully defining polycentric political systems and describing some of their functional properties. In the following section, I present the key feature of our theoretical framework: the complexity problem. As societies increase in scale and diversify their productive activities, governors face increasingly high information-processing demands. This presents a governance dilemma: the governor can maintain direct control only by resisting social complexity. Governing complexity, by contrast, demands multiple, semi-autonomous centers of power. In other words, the increasing informational demands of complexity exert pressure on states to shift from a direct control paradigm to a polycentric paradigm. After explaining the logic of the model, I draw upon two types of empirical evidence. First, in order to test the relationship between polycentric governance and social complexity within a large sample of diverse societies, I rely on the Seshat databank, which aims to quantify and compare social complexity on a global scale throughout history. I show that the hypothesis gains modest support from this large sample. Second, I develop a detailed case study that compares alternative trajectories in the evolution of Western European governance structures. This is important, since even the best cross-cultural historical data available is rather sparse and unsuitable for illustrating the finer details of my account. The selected case study offers a proof of concept and highlights the mechanisms that underlie the explanatory model.
5.2 Polycentric Political Orders

At its most abstract, the concept of polycentricity can describe systems as disparate as scientific inquiry (Tarko 2015; Polanyi 1951), democracy (V. Ostrom 1972), markets (Boettke and Candela 2015), and the common law (Hayek 1973). In all of these realms, a network of connected decision makers seek out solutions to pressing problems. Through an evolutionary process of trial and error, ever better solutions come to replace dysfunctional and outdated ones. The classic analysis of polycentric political order appears in an article by Vincent Ostrom, Charles Tiebout, and Robert Warren, who identify several of its key features. Polycentric political order exhibits “overlapping jurisdictions” of authority, “duplication of functions” with respect to public goods and services, and “many centers of decision making that are formally independent of each other” (Ostrom, Tiebout, and Warren 1961, 831). Such an arrangement supports competition both within and between various levels of organization. To elucidate the concept further, Paul Dragos Aligica and Vlad Tarko have isolated three essential features (2013, 737; 2012, 252-3):¹

1. several centers for decision making that act with relative autonomy within their (possibly) overlapping jurisdictions,
2. a single overarching system of rules (formally or informally enforced), and
3. the emergence of a spontaneous social order as the outcome of an evolutionary competition between different governance units.

Illustrative examples of polycentric political orders include metropolitan governance, American federalism, and the European Union (Vogler 2020). Considering the latter example, decision centers include the municipal governments of European states, the national governments, regional coalitions, as well as the European council, commission, and parliament. The jurisdiction of these decision centers clearly overlap: cities can pass ordinances, even while the laws passed at the highest level of the European government, the overarching system of rules, will also apply to those cities. This overarching system of rules regulates how citizens and governments of various nations interact with one another. One specific feature of this system of rules is that it permits the free movement of goods and people between European nations, allowing for both market competition as well as “Tiebout competition,” i.e. voting with one’s feet, both of which support spontaneous legal and economic orders (Mueller 2003, 187; Tiebout 1957).

One variable related to spontaneous order identified by Aligica and Tarko is of particular relevance for the present argument. Polycentric political orders permit the flow and aggregation of information in a way that monocentric states cannot. As Hayek (1945) famously argued in the economic context, the phenomena of rivalry and competition facilitates the use of vast amounts of tacit, dispersed, and inarticulable information. To find the cheapest methods of production, for example, several firms must engage in small scale experiments, trying out different combinations of inputs and producing slightly different products. The emergent prices, which provide an interface between consumer demand and factor supply, determine which firm has successfully met consumer needs at the least cost. This firm will enjoy profits, while other firms will experience losses, reducing their market share over time unless they self-correct. Without

¹ Stephan, Marshall, and McGinnis (2019, 41) concur with this analysis, though they place special emphasis on feature (1). In their concept design, Aligica and Tarko (2012, 257) identify empirical indicators for each of the three essential features. They ultimately propose ten necessary conditions a polycentric order must satisfy, several of which are exclusive disjunctions. Explaining this highly detailed concept design is unnecessary for present purposes.
the process of experimental entrepreneurship, information about the best and least costly production methods would remain underutilized or even undiscovered.

In the same way, individuals within a political unit possess knowledge that a central government does not. Polycentric political orders permit the use of this information through horizontal and vertical competition. Horizontal competition occurs between political units on the same level, typically those without overlapping geographical regions. Tiebout competition facilitates horizontal competition, and thus provides some information about which laws and policies are working. Vertical competition occurs as citizens opt to organize the provision of governmental services on different scales. For instance, policing services can be administered by the municipality or at the federal level. When democratic choice procedures are in place, different levels will compete for voter support. More generally, democratic institutions, which grant decision-making power to the lowest-level unit, the individual, aim to aggregate information about voters’ preferences and beliefs. This information is harnessed to provide feedback for political leaders who are experimenting with different policy solutions (Barrett 2020). Citizens who are unhappy with a regime can contest it by rallying political support in its opposition. In its ideal form, democracy thus exhibits impressive epistemic properties (Landermore 2017; Landermore and Page 2015; Estlund 1997; Cohen 1986). By incorporating contestation, experimentation, and competition, this polycentric paradigm is able to utilize dispersed and tacit knowledge in a manner analogous to the market-price system.

By contrast, a monocentric state, relying on the direct control paradigm possesses much less information about the conditions “on the ground.” To some extent, this is an issue with any attempt to manage a situation from the top of a hierarchical bureaucracy (Scott 1998: 76). The monocentric state faces especially severe informational limits in this respect; they lack the bottom-up feedback of the democratic process, as well as any process of internal Tiebout competition to signal the success and failure of heterogeneous policy regimes. The direct control paradigm prevents the endogenous emergence of policy decisions, instead attempting to govern in a top-down fashion. The result is informational poverty.

Yet, monocentric administration is not without its advantages. As I will explain in more detail below, the direct control paradigm limits transaction costs by cutting out bargaining stages and by ensuring coordination on well-defined projects. Which governance structure is more effective depends, crucially, on an important variable, social complexity, to which I now turn.

5.3 The Complexity Problem

What is Complexity?

Intuitively, many contemporary societies exhibit unprecedented levels of complexity. To take one illustration, Eric Beinhocker has noted that the number of different goods currently for sale in New York City far exceeds the number of species on the planet. This is not the case, not

2: Some are far less sanguine about the epistemic value of democracy (Brennan 2016). But see Christiano (2015).
3: The two processes outlined in this paragraph—Tiebout competition and democratic choice—align with Hirschman’s (1970) two famous modes of organizational reform: exit and voice.
4: In the framework of Buchanan and Tullock (1962) the “decision-making costs” are minimized in a monocentric political structure.
The Problem of Complexity and the Emergence of Polycentric Political Order

5 The Problem of Complexity and the Emergence of Polycentric Political Order

even remotely so, for simpler societies such as the Yanomani (Beinhocker 2006). Although the types of goods and services exchanged in an economy is a sign of social complexity, it is merely a proxy. What, fundamentally, distinguishes a simpler society from a complex one? Social scientists have addressed this question both conceptually and empirically. The concept of complexity has been roughly characterized by Scott Page (2015) as involving five properties. A system is complex when it comprises:

1. a large numbers of entities,
2. a network structure that links these entities,
3. high levels of interdependence between these entities,
4. adaptive behavior on the part of these agents, and
5. diversity or heterogeneity of these agents.5

In light of these properties, complex systems exhibit emergent properties, nonlinear dynamics, and constant novelty. As a result, traditional modeling and long-term prediction become impossible. Many consequences of intervention in complex systems will be unforeseen yet significant.

The empirical approach to analyzing complexity supports this conceptual analysis. The most extensive empirical appraisal of complexity in historical societies, the Seshat data set, will be more carefully discussed in the next section. For present purposes, it suffices to note that many of the features identified by Page have been associated with empirical indicators such as the total population and the number of different types of administrative officials (Turchin et al. 2020). When large numbers of heterogeneous individuals interact, the Seshat data set suggests, social complexity ensues.

Social Complexity and Direct Control

Social complexity is highly significant when considering the scope and limitations of political actors. To begin our analysis, consider a state that operates in accordance with the direct control paradigm. Such a state has three key features. First, it issues commands unilaterally from higher levels to lower levels of political administration; the direction of command is from the head of the state to its subordinates. The subjects do not issue commands to the state. Second, it actively opposes alternative centers of power, be these other semi-independent governing units or autonomous, private corporations. Finally, power is organized in a vertical fashion, with few horizontal connections.

In a complex society, this sort of state faces severe informational challenges of two types. The first is the bottom-up challenge, alluded to above, namely, that of apprehending vast amounts of dispersed and tacit information. While the second is a top-down challenge, namely, that of predicting the important, long-term effects of policy interventions.

James Scott (1998) expertly analyzes the bottom-up challenge of state administration. “Officials of the modern state,” Scott writes, “are, of necessity, at least one step—and often several steps—removed from the society they are charged with governing. They assess the life of their society by a series of typifications that are always some distance from the full reality these abstractions are meant to capture” (76). Public administrators of states that govern complex societies cannot possibly access,
document, and digest the heterogeneous vernaculars of their myriad subcommunities. They have difficulty perceiving complex patterns that individuals on the ground are highly adept at utilizing. Consequently, administrators must rely upon “typifications,” such as statistical aggregates. Relying on such “synoptic facts” is indispensable to statecraft (77). Yet, doing so comes at a cost:

These state simplifications, like all state simplifications, are always far more static and schematic that the actual social phenomena they presume to typify. The farmer rarely experiences an average crop, an average rainfall, or an average price for his crops. (46)

In other words, each time a complete state of the world is distilled into a simple, abstract statistic for use at a higher level of administration, information is lost. Often, this information is more than a distracting detail. As Scott shows, it is impossible to know which pieces of information will be crucial for achieving desirable outcomes, and which can be safely set aside. This issue is less severe when the governing unit is closer to the subjects being governed, when the subjects being governed are homogeneous, and when these subjects inhabit relatively predictable and uniform environments. However, upon introducing diversity, interdependence, and larger populations — i.e. upon increasing complexity — the loss of information becomes increasingly severe. In short, in the absence of institutional solutions, bottom-up information loss is an increasing function of social complexity. And without adequate information about local conditions, the effectiveness of state interventions becomes highly suspect. Success will be evaluated according to abstract metrics that fail to capture crucially important details about the true state of the world.

The second, top-down informational challenge also threatens to undermine the effectiveness of state interventions into society. When systems become complex, they defy prediction. Simple models assuming either a small number of elements or homogeneity fail to provide insight into the dynamics of complex systems (Weaver 1949). The impossibility of using the experimental method to determine predictive laws for society has been recognized at least since Mill (1836). At its most basic, the impossibility of prediction is a result of the combinatorics involved in systematic attempts to determine predictive laws. As Page has pointed out, even if society were to possess only ten relevant variables that could take on three possible values, social scientists would need to run $5 \times 10^{17}$ separate experiments to determine the effects of changing the values of these variables (2018: 5). If there were a mere 100 relevant variables, we would require $5.15 \times 10^{37}$ separate experiments! Even more radically, Karl Popper (1948/1962, 339) has suggested that predictive laws may simply not exist in social systems. Such systems are technically “open,” that is, subject to influence from various exogenous effects. Because of their interdependence and adaptiveness, agents react to these exogenous effects in ways that cause profound changes in the social system. As a result of complexity, therefore, predictability is greatly limited. Consequently, a state that attempts to impose, in a top-down fashion, a particular outcome will fail, unless, perhaps, that outcome is narrowly defined, so as to ignore all of its unforeseen ramifications (Gaus 2021). A lack of precise predictability impedes control. As a consequence, we
observe, once again, an inverse relationship between state effectiveness and social complexity.

Both of these informational challenges point towards a fundamental problem that political institutions face when confronted by an increasingly complex society.

The Complexity Problem: The ability to understand and control social outcomes, that is, to effectively intervene into society, decreases as social complexity increases.

At some point, the complexity problem becomes so severe that the direct control paradigm is no longer workable. In such a situation, the state faces a choice between two options. First, it can fight against complexity. To do so, it restricts the diversity and spontaneity of actors in its society. Scott notes that all states, to some extent, engage in simplifying practices; they attempt “with varying success to create a terrain and a population with precisely those standardized characteristics that will be easiest to monitor, count, assess, and manage” (81-2). Attempts at social simplification become especially critical for states organized around the direct control paradigm. If they succeed at holding social complexity at bay, then the direct control paradigm may remain workable. Otherwise, they must opt for the second option: shifting to a new governance paradigm. The governance paradigm that typifies most modern states is what I call the polycentric paradigm. Like the direct control paradigm, the polycentric paradigm is an ideal type, approximated to different extents by various contemporary states. We now turn to the characterization and examination of this paradigm.

Social Complexity and Polycentricity

The polycentric state differs from the direct control state in several key ways. First, while direct controls may sometimes rely on hierarchical levels, this is a necessary prerequisite for the polycentric state, since polycentric governance requires that services be provided on multiple scales (Stephan et. al. 2019, 41; Aligica and Tarko 2012, 257; V. Ostrom 1972). While a direct control state may establish a hierarchy with unidirectional authority (lower levels answer to higher ones), polycentricity requires that these layers exhibit some degree of autonomy and that lower levels exert some influence on higher levels. The subjects, mid-level functionaries, and high public officials answer to one another.

Second, the polycentric state permits competition between administrative layers (V. Ostrom 1972). If local efforts fail to adequately defend the local populace, defence services can be organized at a higher level, such as a confederation of cities, or a nation. This may be undertaken through a “quasimarket” arrangement, where local governance structures purchase such services from lower cost providers who reap the advantages of economies of scale (Boettke, Coyne, Leeson 2011).

Third, the polycentric state supports alternative centers of power within each scale, be these other governing units or semi-autonomous, private organizations. For instance, in the United States, a national government provides services to state governments. In the European Union, the European government supports national governments in various ways. These

---

7: One way of resisting complexity is to break apart into smaller political units once population increases beyond a certain point. This was a characteristic response of feudal manors around the 11th century (North 1981, 132).
The Problem of Complexity and the Emergence of Polycentric Political Order

One important cost involves competition among society’s elites, which can become especially severe in polycentric orders. For an example of this in the highly polycentric Italian city states of the 15th century, see Chittolini (1989).

Historically, of course, few if any societies have been democratic in the sense of holding elections or referenda that include a large percentage of the population. However, polycentric political structures need not be democratic in this sense. Polycentricity as I have defined it simply requires that lower levels of political structure exert power or influence on higher levels.

Polycentric governance involves a larger, less streamlined government. The various nodes of power must negotiate before many resolutions can pass. As ever more agents acquire votes or veto rights, reform becomes much more difficult; “[a] political system that is all checks and balances is potentially no more successful than one with no checks” (Fukuyama 2011, 431). Direct control avoids the costs associated with political bargaining and coordination, since a single individual or small group determines the social choice, simply handing it down without public discussion or deliberation. Aside from the costs associated with political decision making, the diversity of rules and institutions that polycentricity supports also impose a cost. Suppose, for example, that contract law differs between jurisdictions A and B. The need to research and negotiate the correct contractual form will introduce an additional cost that would not exist in contract law was uniform. Polycentricity thus introduces additional transaction costs in both political and economic spheres. Direct control, in other words, minimizes transaction costs when compared to polycentricity.

Nevertheless, informational problems come to the fore as social complexity increases. As we saw, in the direct control paradigm detailed information on the ground struggles to influence policy issued by high public officials. Polycentricity alleviates this problem in various ways. First, much governance is undertaken by lower-level political units. Because these units have smaller constituencies as well as more lines of communication between their officials and their citizens, the officials and citizens enjoy much greater proximity. Second, insofar lower administrative levels can affect higher ones, the authority between various levels is bidirectional, allowing citizens or small organizations to directly and indirectly register their preferences and beliefs. For instance, in the United States, citizens at the lowest level of organization select the president directly, but they also select senators, representatives, and governors who influence the president in various ways. This provides a bottom-up source of information.

Competition is also paramount to generating usable information. Vertical competition occurs when citizens decide whether local, regional, or federal governments best provide public goods and services. This information guides the state in determining the best level at which to provide a public good or service. Horizontal competition occurs as citizens relocate to more effective polities. Due to transaction costs, horizontal competition is typically less effective than market competition. Nevertheless, when a governing unit is sufficiently ineffective, citizens will emigrate or, at least, decide not to immigrate. Horizontal competition thus provides a

states (in the U.S.) or nations (in the E.U.) offer alternative governance schemes to a population of citizens who select their jurisdiction partially on the basis of its governance competency.

In sum, polycentric governance entails a multitude of levels, and competition occurs both vertically and horizontally.

In situations of low complexity, it is not clear that the polycentric paradigm has any advantages over the direct control paradigm. If citizens are homogeneous in their values/preferences, low in number, relatively autarkic, and reside in stable, uniform environments, then the informational problems noted above will be minimal. In fact, in such a scenario, the polycentric paradigm seems to engender unnecessary costs. Polycentric governance involves a larger, less streamlined government. The various nodes of power must negotiate before many resolutions can pass. As ever more agents acquire votes or veto rights, reform becomes much more difficult; “[a] political system that is all checks and balances is potentially no more successful than one with no checks” (Fukuyama 2011, 431). Direct control avoids the costs associated with political bargaining and coordination, since a single individual or small group determines the social choice, simply handing it down without public discussion or deliberation. Aside from the costs associated with political decision making, the diversity of rules and institutions that polycentricity supports also impose a cost. Suppose, for example, that contract law differs between jurisdictions A and B. The need to research and negotiate the correct contractual form will introduce an additional cost that would not exist in contract law was uniform. Polycentricity thus introduces additional transaction costs in both political and economic spheres. Direct control, in other words, minimizes transaction costs when compared to polycentricity.

Nevertheless, informational problems come to the fore as social complexity increases. As we saw, in the direct control paradigm detailed information on the ground struggles to influence policy issued by high public officials. Polycentricity alleviates this problem in various ways. First, much governance is undertaken by lower-level political units. Because these units have smaller constituencies as well as more lines of communication between their officials and their citizens, the officials and citizens enjoy much greater proximity. Second, insofar lower administrative levels can affect higher ones, the authority between various levels is bidirectional, allowing citizens or small organizations to directly and indirectly register their preferences and beliefs. For instance, in the United States, citizens at the lowest level of organization select the president directly, but they also select senators, representatives, and governors who influence the president in various ways. This provides a bottom-up source of information.

Competition is also paramount to generating usable information. Vertical competition occurs when citizens decide whether local, regional, or federal governments best provide public goods and services. This information guides the state in determining the best level at which to provide a public good or service. Horizontal competition occurs as citizens relocate to more effective polities. Due to transaction costs, horizontal competition is typically less effective than market competition. Nevertheless, when a governing unit is sufficiently ineffective, citizens will emigrate or, at least, decide not to immigrate. Horizontal competition thus provides a
political “quasi-market,” in which the exercise of exit power permits the flow of information from citizens on the ground to higher-level political units (V. Ostrom 1972).

At the same time, polycentricity can be viewed as a partial solution to the top-down information problem. Most obviously, shifting governance responsibilities to lower level units will reduce the average complexity of the problems that a government must solve. When lower level units call the shots, constituencies will be smaller and more homogeneous, making policy interventions more predictable, while collective goals become better defined and more achievable. In addition, the processes of competition provide informational feedback by directing power and resources towards political units that engage in successful experimentation. Political units that stagnate or engage in unsuccessful experimentation will, correspondingly, shrink in relative size and power. As a result, even if the effect of policy interventions is impossible to predict ex ante, the process of competition will ensure that ex post successful interventions will be scaled up, while unsuccessful policies will diminish in prevalence. Both vertical and horizontal competition thus contribute to an evolutionary search algorithm that identifies effective reforms.

A Simple Model of Governing Complexity

In light of the difficulties faced by direct control and the advantages of the polycentric paradigm in coping with increasing complexity, we can posit a rough mathematical characterization of government effectiveness (figure 1). Direct control (blue curve), I have argued, is a low cost technology for governing low-complexity societies, at least with respect to gathering and utilizing pertinent information. Accordingly, the marginal value of governance is quite high when social complexity is low.\textsuperscript{10} However, decreasing returns to scale set in as social complexity increases. Perhaps the strains of complexity are not felt at first, but beyond some point, the function will be strictly decreasing. By contrast, the polycentric paradigm (red curve) generates unnecessary transaction costs at low levels of complexity. The elaborate processes of competition, preference aggregation, and service duplication are wasteful in the absence of informational challenges. However, polycentric governance, largely due to its decentralized structure, scales far more gracefully than direct control. As complexity increases, polycentric governance is able to access and utilize ever more information at a relatively constant marginal cost.\textsuperscript{11} Eventually, there comes a threshold of complexity at which polycentric governance yields higher marginal benefits for the society it governs. That is, it better enables relevant information to determine decision making.

Beyond a certain point of complexity, a direct control paradigm may stifle information flows, rather than help to gather and utilize them. This occurs at the point $c_1$, where the blue curve intersects with the abscissa. There also exists a point where governance itself, including a polycentric approach, may simply become difficult or impossible (Gaus 2021), so even the polycentric governance curve will likely decrease beyond a certain point.\textsuperscript{12} However, we restrict ourselves to intermediate levels of complexity, where the effectiveness of polycentric governance can be represented by an increasing function of social complexity.
Given these two functional forms, there will exist a critical threshold of complexity, $c^*$, at which point the polycentric paradigm becomes a more effective governance form than direct control. This is the point at which the complexity problem becomes acute: the state must either embrace polycentricity or combat the trend of increasing complexity.

**Implications of the Model**

This formalistic description is not a historical narrative; societies do not inevitably progress from less complex to more complex or from monocentric to polycentric. For one thing, the chronology may be quite different than described above. Rather than first arising as a direct control state and shifting to a polycentric state in the face of increasing complexity, some modern states have been polycentric from their conception.
According to John Hall (1985), this is the correct way to think of the "organic" European state. Another important point, discussed in more detail below, is that the shift from monocentric to polycentric need not be conscious or deliberate. I have described it in that way for ease of expression, but the history of political development is one of accident and contingency, not of foresight and intention.

More to the point, the marginal values of the two respective paradigms are the marginal values to society, which is not necessarily the marginal value to those making decisions for society. If we assume that the ruler is revenue-maximizing, then there is no reason to think that what promotes general economic prosperity is necessarily what will promote the ruler’s interest, especially in the foreseeable short-term. Adopting Douglass North’s model, we might think of the ruler as a monopolist who distributes services at an inefficiently high price (North, ch. 3).

In this case, giving up monopolistic power would be socially beneficial, but not politically desirable. Consequently, direct control states may be stable, even if they are inefficient. Despite increasing social complexity, the ruler may refuse to decentralize her political power. Of course, given that there are gains to be had, if a profit sharing scheme could be devised and enforced, then the ruler would gladly decentralize political power. However, transaction costs are generally too high for this sort of contractual arrangement; by the very nature of relinquishing political power, such a contract would be unenforceable. So society might remain stuck with ineffective governance institutions despite the fact that its high level of complexity calls for polycentric institutions. This can be thought of as a high-complexity, low-polycentricity equilibrium.

This kind of equilibrium appears to be relatively unstable, since monocentric states are incapable of accurately monitoring or producing desired outcomes in highly complex societies. Even a task as simple as extracting taxes has proven immensely difficult for central authorities throughout history. As we will see in the case studies that come later, states often responded to this difficulty by striking deals with local guilds in order to facilitate the measuring and taxing of output. This, however, had the effect of simplifying society by reducing the number and variety of producers in the economy. The important point here is that social complexity is not a dependent variable. In an effort to more easily extract taxes, corvée labor or other resources, states strive to render society more “legible;” they can and have undertaken what James Scott calls “heroic simplifications” (Scott, 2, 8, 63). In this sense, societies may often remain or arrive at a low-complexity, low polycentricity equilibrium.

Hampering complexity, in general, involves limiting the diversity and interdependence of one’s citizens. This poses a limit to economic prosperity, based as it is on a deep division of labor and widespread production and exchange activities undertaken by individuals with access to capital not controlled by the state. So, just as maintaining monocentricity in a complex society is detrimental to economic prosperity, so limiting complexity in order to maintain effective control also results in large welfare costs. These observations allow us to tease out several predictions of the model.

First, when the right sort of competition is present, societies will tend towards complexity and states will tend toward polycentricity. In this case,
The Problem of Complexity and the Emergence of Polycentric Political Order

In a provocative discussion of capital mobility and competition in early modern Europe, Hirschman (1978) suggests that “understood complexity” may itself be a public good of sorts. It might, therefore, be a dimension along which states compete.

Second, the wrong sort of competition might yield the opposite result. If competition occurs along militaristic lines, rather than cultural or economic, then a monocentric state may enjoy an advantage. Although information about local preferences, resources, and production possibilities will be lost in a monocentric political structure, these bits of information are less important when a single goal — e.g., military hegemony — is pursued. In that case, society will likely remain at or convert into a low complexity, monocentric equilibrium. As Hayek (2014), among others, has noted, a so-called “war-time economy” is technically simpler than a free economy in a state of normalcy.

Finally, in a state of low competition, the rulers acquire greater monopoly power and, consequently, the revenue curve of the rulers will deviate substantially from that of society at large. In this case, the low complexity, monocentric equilibrium is the predicted outcome. Indeed, some have attributed China’s comparative decline after the 15th century to a lack of competition between political units (Ferguson, 1998-99).

In sum, the theory does not predict that all political units will increase in complexity and move towards polycentric governance structures. Instead, the analysis suggests that polycentric states better accommodate social complexity, while direct control states will tend to wither upon its appearance, or effectively stunt social complexity altogether. This latter phenomenon is how Hall (1985) conceives of the Chinese empire, which actively sought to cripple economic and technological developments that might undermine the power of the “capstone” state. The direct control state, in other words, must strive to maintain a level of complexity below \( c^* \) lest their power should prove ineffectual in governing their territory.

Although some of the language used here has made it sound as if rulers consciously select between monocentric and polycentric political institutions, I must re-emphasize that this is not entailed by the model. Although deliberate design is possible, the case studies examined below show that political institutions often stumble along blindly as rulers attempt to adapt to the local pressures of time and place. These pressures may impel them to decentralize their authority, but, at least in the cases we examine, there is no conscious intention to develop polycentric political institutions. The interests and foresight of rulers are typically far too narrow for that. Before exploring detailed cases, however, it is worth asking whether this hypothesis is supported by any large-scale historical data. The next section will examine the Seshat data set and subject the theory outlined here to some preliminary empirical tests.

5.4 Empirical Analysis

The proposition under consideration is that, at least beyond a certain point, increasing social complexity is incompatible with centralized control.
Empirically, this means that there should exist a strong correlation between growing social complexity and the emergence of polycentricity. Although I have carefully explained the meanings of these terms, defining empirical indicators for such abstract notions is itself a difficult task. In defining such indicators, one faces two competing desiderata. First, there is the conceptual question of precision: does the empirical indicator truly measure the relevant concept? Second, there is the practical question of data acquisition: can we collect data on the empirical indicator from a sufficiently large sample to do a proper hypothesis test? Although for some societies we possess reliable and detailed information about social and governmental structures, to collect a sufficiently large sample we must rely upon rough proxies, rather than the richer notions of complexity and polycentricity explained above. This is necessary since, in order to scientifically test propositions, that is, in order to determine whether these propositions are likely to generalize to societies other than those we study, it is imperative to consider a large sample of different historical societies. We thus face a trade-off between precision and generality.

In order to adequately address this trade-off, I provide both kinds of empirical evidence.

**Social Development and the Seshat Data Set**

Peter Turchin and several co-authors have recently collected large-scale data on social complexity across a vast temporal and geographical scale. The resulting data set, known as the “Seshat” data bank,(Turchin et al. 2015) examines over four hundred societies from thirty regions and spans over 10,000 years. The various measurements taken from these societies are divided into nine categories, called “complexity characteristics.”

Some of these complexity characteristics correspond roughly to aspects of polycentric governance structures, as defined above. Others correspond more closely to social complexity proper. Crucially, empirical support for the simple model developed above comes from subsequent analyses of relationships between these various complexity components; these analyses reveal that these complexity characteristics are tightly correlated and that they “coevolve in predictable ways” (Turchin et. al. 2017, E144).

To be more specific, the key indicators corresponding to social complexity, as defined above, include polity population (CC1), polity territory (CC2), the size of the largest settlement (CC3), as well as information and communication technologies (CC7 and CC8). These complexity characteristics indicate larger numbers of people in closer contact with one another, the essence of social complexity.

The indicators for polycentric governance are rougher than those for social complexity. Ideally, the indicators would measure (1) hierarchy, (2) decentralization, and (3) competition. The Seshat databank provides excellent indicators for (1), very rough indicators for (2), and no indicator for (3). The complexity characteristic (CC4) looks at hierarchy specifically, thus providing a measure of (1). (CC5) looks at the diversity of government officials and institutions. For example, if there are priests involved in governance, not just a King, then (CC5) receives a higher score. Similarly, if there are independent courts and a legal code, then (CC5) will also
score higher. Because a diversity of government officials and higher numbers of government agencies represents a delegation of authority, (CC5) should correlate with polycentric governance.

Of course, these indicators, especially for (2), are flawed. Highly monocentric governments might have elaborate bureaucracies that score high on (CC5). However, a monocentric government might also score low on (CC5), while a polycentric government is less likely to do so. Polycentric governance requires delegation of authority in its very definition, while monocentric governance is merely compatible with such a delegation. Analyzing this data can, nevertheless, provide a preliminary test, but nothing more, for the hypothesis that social complexity requires polycentric governance. The potential to falsify this hypothesis is very real, even while a failure to reject it provides no solid evidence of its truth. This Popperian approach must suffice until better data is acquired, but it will be supplemented by the detailed discussion of historical cases in the next subsection.

In their analysis of the Seshat data, Turchin et. al. determine that there exist “functional relationships between these characteristics that cause them to coevolve,” just as the model would predict (Turchin et. al. 2017, E147). Moreover, the authors speculate that as complexity increases, autocratic forms of power are displaced by collective ones (E149). Support for this comes from the observation that collective forms of power seem to emerge later and in more complex societies than do autocratic forms of power (Turchin et al. 2015). Collective forms of power are not exactly synonymous with polycentricity, but the move away from autocratic control and towards more dispersed forms of power certainly resonates with the hypothesis under consideration.

In sum, the Seshat data set, currently the most advanced and complete data set measuring social complexity, offers modest support for the hypothesis under consideration. At the very least, it does not falsify this hypothesis, though it very well could have. It certainly could have shown that increasing complexity calls for more centralization and fewer hierarchical levels. This would have falsified the proposed theory. Yet, the data set shows the opposite. Why should this be the case? I have proposed an answer to this question. To see if this answer is compelling, it is also beneficial to zoom in on some detailed, historical cases.

**Emerging from Feudalism**

To achieve a more fine-grained image of the relationship between social complexity and polycentric governance, the transition from feudalism to the nation state in Western Europe provides a striking case study. Although all European nation states eventually transitioned to liberal, market-based economies, they didn’t do so at a uniform pace. In England and the low countries, trade and industry took off early, generating a rapid increase in complexity and prosperity. In France and Spain, by contrast, monocentric governance went hand in hand with dysfunctional markets and slow growth, resulting in Malthusian checks on continued growth. A third case is that of the non-feudal confederations of city states present in Italy and Germany. Initially polycentric, their capacity to maintain a decentralized political organization slowly eroded, largely disappearing
The Problem of Complexity and the Emergence of Polycentric Political Order

by the 17th century. Examining each of these three cases illuminates the causal connection between complexity and polycentricity.

The conditions characterizing European political organization prior to the 10th century A.D., taken here as the initial conditions, were somewhat similar across our first two regions of interest. In England, France, Holland, and Spain, the collapse of the Roman empire in the fifth century led to political anarchy and a general absence of law and security. Germanic hordes made a living on raiding and pillaging, leading to the form of economic organization known as manorialism. This organizational form involved a quasi-contract between serfs and lords (North and Thomas 1971, 778). According to this contract, serfs owed their local lord a certain amount of labor input, roughly three days a week in many places, in exchange for protective services from the lord. Due to the dangers of traveling beyond the manor, trade was limited; manors were more or less autarkic. Land was abundant and therefore cheap, while labor was competitively sought by feudal lords. Initial economies of scale were quickly swamped by increasing marginal costs of protection, resulting in the U-shaped marginal cost curve much beloved by teachers of microeconomics (North 1981, 126-9). In other words, the value of protective services diminished appreciably as farmland expanded past a certain distance from the center, and manors could not, therefore, grow beyond a certain size.

Before charting the evolutionary courses that followed from these initial conditions, a conceptual point bears emphasis. While highly decentralized, the governance structure in Europe around 900 A.D. was not polycentric. Decentralization alone does not suffice to make a system polycentric. As noted above, polycentricity requires the additional features of interaction and hierarchy. A set of autarkic, monocentric orders does not become polycentric until (1) they begin to interact and coordinate their behavior under an overarching set of rules, and (2) governance is offered on a multiplicity of scales (Tarko forthcoming). The political and economic relationships between the manors were negligible. And while the feudal system did involve kings who technically operated at a higher governance scale, their power in the 10th century was trivial. To become polycentric, the initial system would require greater integration and stratification.

This arrangement was an effective response to the chaos that engulfed Western Europe. And, just as importantly, it could scale. Due to the existence of ample uninhabited territory, when manors expanded past their most efficient size, new lords could start new manors. As ecological theory would predict, population steadily increased and continued to do so until a host of Malthusian checks (mostly plagues and famines) reversed this trend in the 14th century. At the same time, changing technology also placed new stresses on traditional forms of political organization. Warfare, in particular, became far more advanced and required far more revenue to carry on competitively. Gunpowder made feudal fortifications increasingly irrelevant. Against this changing background, feudal lords, the Catholic church, and kings all jockeyed for power.

As Douglass North has argued in various works, these fluctuations in population and technology had profound effects on relative factor prices, which, in turn, had profound effects on contractual arrangements
and, ultimately, on the basic institutional framework of Western society. Throughout all of Europe, the growth in population led first to increased gains from trade, especially as the frontier of European society expanded to ever more diverse geographies. Secondly, partially as a result of increased gains from trade, larger political units, capable of providing security along trade routes, became increasingly advantageous. In this way, North and Thomas (1971) explain the increased prominence of larger political units, headed by monarchs. The institutional reforms sparked by changes in population were not, however, homogenous, but exhibited some striking differences, differences which will illuminate the topic at hand: polycentric governance and its relationship to social complexity.

England and the Netherlands developed comparatively polycentric governance institutions due to a host of factors that limited the centralizing ambitions of monarchical rulers. North (1971, 141, 156) cites three factors, in particular, that limited the ability of English monarchs to achieve complete authority. First, the potential gains that a central government could offer to English inhabitants were comparatively modest. As a relatively secure island nation, England did not require a large and expensive standing army like that of France or Spain. Second, the possession of royal authority was constantly in dispute; alternative rulers were ready at all times to make a play for the throne should its current occupant prove incompetent or unpopular. Finally, given the export-based structure of the English economy, taxation was far simpler. Wool exports are easy to measure and provide an easy source of fiscal revenue. Consequently, the large tax-collecting bureaucracy required by most other monarchs had no place or purpose in England. The English monarch could access tax revenues without constructing a large and invasive state. “In short,” North tells us, “little reason existed to concentrate authority in the crown over property rights and taxation, and still less reason existed to support a large central government” (1971, 156).

These three features contributed to a single outcome: the English ruler acquired less power than the rulers of most other nations. As a result, when they sought to empower English monarchy, the Tudors became beholden to the merchant class for political support in the struggle against the church and nobility. This involved relying upon and legitimizing the House of Commons, as well as granting cities independent charters and laws. English governance thus took on a form that promoted information flow from lower level interests and concerns into Parliament and, ultimately, to the monarch. The authority to set taxes and raise an army eventually fell into the hands of Parliament, meaning that the English government was far more divided and decentralized than that of their chief rivals.

The polycentric nature of the resulting legal order is explained well by Maitland in his constitutional history of England (Maitland 1909, 2001). The king of England had regular interaction with county courts. Counties were governed in a broadly participatory fashion, with all free landholders having a say, regardless of rank. Counties, in turn, were made up of even smaller governance units called “hundreds,” with their own courts and assemblies. The hundreds even administered certain public goods, like policing and punishment. In discussing the representation of counties in Parliament, Maitland tells us...
When in the middle of the thirteenth century we find elected representatives called to form part of the national assembly, of a common council of the realm, or parliament, they are the representatives of the county courts. They are not the representatives of unorganized collections of men, they are the representatives, we might almost say, of corporations. The whole county is in theory represented by its court... The king’s itinerant justices from time to time visit the counties; the whole county (totus comitatus), i.e. the body of freeholders, stands before them; it declares what the county has been doing since the last visitation; the county can give judgment; the county can give testimony; the county can be punished by fines and amercements when the county has done wrong.

(43)

There thus existed three levels—the royal, county, and hundred levels—each of which had its own courts and administration. If a dispute went unresolved at one level, it could often be appealed at a higher one, permitting vertical competition, a hallmark of polycentricity.

Similar conditions supported Dutch decentralization. Due to unique geographical features and exceptional wealth obtained from their commerce-based economy, as well as an auspicious alliance with a strong Protestant power, the Netherlands enjoyed a secure position vis-a-vis foreign monarchs (van Creveld 1999, 117). Because of this, they were able to operate as a confederacy of provinces, eschewing a strong centralized power. As a result of this local responsiveness, the property rights that emerged in both England and the Netherlands enabled trade to flourish. They thus proved far more efficient than those that emerged in the more centralized regimes of Spain and France.

In both Spain and France, the benefits of centralized national governance were higher and the opportunity costs were lower. The monarch, in other words, had greater bargaining power, as well as a greater need to establish centralized control. Difficulty in taxing local economies led to large bureaucracies and the sale of property rights to the highest bidder. This typically involved selling monopoly rights to certain producers, sometimes entire towns, in an effort to obtain the revenues required to fund large-scale armies. The monopolistic guild system which developed stifled competition and economic growth. In France, the transition began in earnest following the Hundred Years War. In order to monitor and effectively tax regional economies, the crown required the cooperation of local organizations, hence the alliance with local guilds. By the time of Louis XIV, France had achieved full centralization and funded an expensive state bureaucracy largely through the sale of monopoly privileges and venal offices. Local officials were replaced with royal officials known as intendants, who greatly reduced the level of local autonomy much to the chagrin of French towns.

The Spanish equivalent of the intendants were the corregidores. Echoing the French experience, Martin van Creveld recounts how the Spanish monarchy of the 15th century slowly encroached upon local governance institutions, ultimately imposing a fully top-down system:

The Catholic kings were anxious to whittle away the towns’ independence... Already since the fourteenth century occa-
The Problem of Complexity and the Emergence of Polycentric Political Order

A third and final instructive case is that of the loose confederations of city states that emerged in Italy and Germany. From the 14th to the 16th centuries, highly polycentric governance structures pervaded Northern Italy. The system was based on a federal order of towns, republican liberties within the towns, and strongly held ideals of civic life. This system proved robust to centralizing attempts from within the Italian peninsula. During the time in which this governance structure remained intact, Italy became the dominant cultural and commercial center of Europe. Indeed, it was during this time that the Italian Renaissance blossomed into one of the most fruitful cultural epochs in the history of human civilization.

At the same time, several German cities also enjoyed a substantial degree of autonomy. Some of these German city states merged into a confederacy known as The Hanseatic League, which came to dominate maritime trade in the Baltic sea for over a hundred years. Although not all towns in the league were self-governing, several of them were. Yet, they voluntarily cooperated under the governance structure provided by the league.

Both the Italian and German city states ultimately succumbed to centralizing forces. In Italy, these forces were largely external. Despite their commercial and cultural success, the Italian cities were at a militaristic disadvantage relative to the centralized states of Spain and France. Rather

sional corregidores had been sent to some Castilian towns to oversee their affairs — then as now, the phrase “I am the from the government; I am here to help you” was one of the greatest lies in any language. In 1480 it was decided to introduce them into all towns which did not yet have them and to make their office permanent ... the corregidores also acquired administrative authority. (van Creveld 1999, 113)

Despite a series of revolts in the 16th century, the Spanish crown prevailed in wresting power from local governments.

As in France, a catastrophic consequence of top-down rule in Spain was the obstruction of efficient property rights. Whereas tax policies in England and the Netherlands were influenced by the interests of merchants, manufacturers, and landowners on the ground, in Spain the only information and interests that held sway were those possessed by the crown. Given the immediate objective of the monarchy — viz. military hegemony — the crown sought the most expedient means of maximizing state revenues. Again, as in France, this involved granting exclusive monopolies to town-based guilds. More unique to Spain, and perhaps more economically damaging, were the privileges granted to shepherders or mestas. These particular privileges, which had a long history in Spain, guaranteed to the mestas the right to graze sheep back and forth across Spain. Consequently, normal property rights did not develop; land that could be profitably enclosed remained in the commons. Granting such special privileges did, nevertheless, serve the interests of the crown, at least in the short run, by quickly raising substantial revenues. As Daron Acemoglu and James Robinson, among others, have argued, the “extractive” institutions of Spain were transplanted to the New World, where they have ever since hampered economic growth (Acemoglu and Robinson 2012; Acemoglu, Johnson, and Robinson 2002; Fukuyama 2011).

20: For a detailed account of the Italian city states, see Chittolini (1989). For a discussion of the German Hanse, see van Creveld (1999, 109-10) and Johanek (2010).
than relying upon disciplined national armies, they used their immense wealth to hire mercenaries for protection. According to van Creveld,

[...]he results of this policy became apparent in the years after 1494. In terms of economic and cultural accomplishment Italy led the world; however, neither quantitatively nor qualitatively could Italian armies match the much stronger foreign ones that invaded the peninsula and fought each other on its territory. (van Creveld 1999, 108)

The Republic of Florence, for example, came to an end following the Spanish siege in 1530.

In Germany, the main centralizing force was internal. As the Hanseatic league lost its monopoly in the Baltic Sea and faced shrinking profit margins as a result of Dutch competition, German princes made an opportunistic power grab. Consequently, “by 1500 there were no free towns left in Brandenburg; in time the system was extended throughout Prussia” (van Creveld 1999, 109).

Three cases have been compared: the relatively polycentric states of England and Holland, the monocentric states of France and Spain, and the highly polycentric system of city states that prevailed for a time in Italy and Germany. Through this comparison, at least three lessons can be gleaned. First, none of the political systems we have surveyed were consciously designed. Rather, they emerged as a rational response by rulers to shifting environmental conditions. In the Netherlands and England, rulers promoted their own power and revenues by sharing it with a broader class of society. Only by empowering Parliament and defining efficient property rights to aid the burgeoning commercial society could the English kings prevail in their struggle against Church and nobility.

Polycentric political institutions were not designed; they were, to borrow a phrase from Hayek, “stumbled upon” (Hayek 1945). The same goes for monocentric institutions like those of Spain and France. The need to extract revenues from complex local exchanges, paired with the greater need for an expansive military, led to the creation of large bureaucracies and inefficient systems of property rights. It was not the intention of the rulers to stunt the development of social complexity any more than it was the intention of the King of England to promote it. Myopic reactions to contingent local circumstances produced different institutional forms with drastically different social outcomes.

Second, as predicted by the model laid out above, increasing complexity went hand in hand with polycentric political institutions. As Western Europe emerged from feudalism, both population and interdependence rose precipitously. The model predicts one of two outcomes: either governance will become polycentric or the state will suppress social complexity to maintain the effectiveness of its authority. In England and the Netherlands, specialization and interdependence exploded under the auspices of polycentric governance. Moreover, assuming the model is credible, it is not a coincidence that it is in these two locations that the industrial revolution occurred most early and dramatically. In Spain and France, by contrast, economic development and regional heterogeneity were actively stifled by rulers. To maintain control from the center,
both regimes effectively simplified the economy by granting sweeping monopolies to particular guilds or towns. This reduced the number of producers in a given industry, undermining the benefits of competition, but greatly facilitated the task of extracting revenues. Put simply, “[t]he benefits of improving the efficiency of markets were . . . sacrificed to the fiscal needs of the state. As a consequence, France [and Spain] did not escape the Malthusian crisis of the seventeenth century” (North 1981, 150).

The third and final lesson is that, even if polycentric political structures excel along cultural and commercial dimensions, monocentric states often enjoy a sharp militaristic advantage. 15-16th century Italy and Germany flourished along cultural and commercial dimensions. Nevertheless, they were overtaken by far simpler and less impressive rivals. One explanation that suggests itself is that monocentric governance structures specialize in diverting resources towards narrowly-defined goals, such as military conquest. Monocentric governance structures, in other words, are better at being monomaniacal. By contrast, polycentric governance structures allow a variety of jurisdictions to pursue a variety of ends.

To pursue a single end requires a costly process to ensure coordination and compliance. Imagine a monocentrically governed nation with a GDP of $1,000,000,000, and a far more prosperous polycentrically governed nation with a GDP of $10,000,000. Assume that the tax rates in both societies are equal, say 10%. If the ruler of a monocentric polity decides to, he can quickly allocate the full tax revenues ($100,000) to purposes of conquest. The polycentric order, however, grants localities some authority in how their funds are spent. Convincing every single local governor to forgo schools, roads, and police in order to fight a war is a hard sell. The transaction costs are immense. A desire on the part of centralizing rulers to reduce the transaction costs of pursuing their personal goals does much to explain their constant drive to reduce complexity and to minimize local autonomy through efforts so elegantly described by James Scott (1999).

The simple model laid out above thus helps us to understand the evolutionary paths taken by various nations in Western Europe as they emerged from feudalism and formed modern states. At the same time, these detailed cases elucidate some of the dynamics of the model. The main thrust is clear: governing complexity in a top-down, centralized fashion proves immensely difficult. State formation took different paths in different places due to the immediate interests of the ruling class. In some places, the pursuit of these interests laid the groundwork for a flourishing of complexity. In others, the pursuit of these interests led to monocentric structures that actively suppressed complexity in order to maintain a steady revenue stream to the crown. In still others, polycentric orders were undermined by monocentric ones which enjoyed a distinct advantage in pursuing the narrowly defined goal of territorial expansion. All in all, these empirical cases support and illuminate the relationship between complexity and polycentricity conveyed by the simple model proposed above.
5.5 Conclusion

In order to provide effective governance in complex societies, I have argued, requires decentralized, hierarchical, and competitive political institutions — i.e. polycentricity. Reasoning in terms of information and transaction costs produces a simple model in which, beyond a certain point of complexity, polycentric political organization is imperative. The relationship between polycentric governance and social complexity posited by this model finds support in large-scale statistical correlations between the complexity characteristics identified by the Seshat database. Additional support and elucidation arises from an examination of increasing complexity in Western Europe following the decline of feudalism. While many factors likely affect the form that political institutions take, empirical evidence supports the idea that polycentric governance structures are a natural response to complexity, or, equivalently, that efforts at simplification are a natural response by monocentric states to increasing complexity.

The difficulty of aggregating and utilizing dispersed information creates a problem for those who wish to govern a complex society. This complexity problem is the political equivalent of what Hayek called “the economic problem of society,” viz. the problem of using tacit and dispersed knowledge to generate a rational economic order. While Hayek argues that the solution to this economic problem is to forgo political action in favor of spontaneous market forces, the thrust of this paper has been that there are ways of structuring political institutions that mitigate the complexity problem. A central planner may be a clumsy way of administering economic policies, but local governments, engaged in various types of competition with other jurisdictions, may not face the same informational limitations that worried Hayek. Rivalrous competition generates information in markets; in the same way, it might generate information in politics.
Dynamic Justice
In this paper, the term "equilibrium" is used to refer to fixed point equilibria. It does not include cycles or any other type of pattern that some might refer to as an equilibrium.

Is Justice a Fixed Point?

An equilibrium centered view is essentially static and provides little insight into the transient behavior of systems that are not near the equilibrium. Natural, undisturbed systems are likely to be continually in a transient state; they will be equally so under the influence of man.


6.1 Introduction

“It is evident,” John Rawls tells us, “that stability is a desirable feature of moral conceptions. . . However attractive a conception of justice might be on other grounds, it is seriously defective if the principles of moral psychology are such that it fails to engender in human beings the requisite desire to act upon it” (1999, 398). Rawls has inspired legions of political theorists to fixate on the comparative stability of different conceptions of justice (Chung 2020; Thrasher and Vallier 2018, 2015; Kogelmann and Stich 2015; Weithman 2017). Within this discussion, theorists typically examine different justice equilibria, that is, states in which a particular conception of justice “generate[s] its own support” over time (Rawls 1999, 138). To test stability, they search for subversive tendencies or imagine subjecting society to diverse shocks of various sizes (Chung 2019; Vallier 2017).

This perspective on justice, however, makes nontrivial assumptions about the nature of social-moral systems. To examine the properties of an equilibrium state, such as stability, one must first justify the assumption of an equilibrium state. Before discussing stability, the theorist must discuss existence.

Showing that an equilibrium exists in a complex system like society can be difficult. Some such systems simply do not exhibit equilibrating tendencies. In fact, there is good reason to suppose that the dynamism and flux of contemporary society are the result of endogenous disequilibrium. This has been forcefully argued in the case of economic systems (Arthur 2015; Schumpeter 2008), as well as other social systems (Miller and Page 2007). It is reasonable, therefore, to suspect that social-moral systems may also resist equilibration. If this is the case, then to accurately characterize the nature of justice, political theory may need to refocus its gaze. As Holling explains in the epigraph, applying equilibrium analysis to nonequilibrium phenomena can blind the theorist to important features of the phenomena being examined.

In this paper I argue that philosophers have granted undue attention to just states of affairs, thus ignoring the process by which such states arise and are swept away. Rawls’s earlier theory of justice presents a paradigm, helping to focus this abstract discussion. The relevant aspects of his
theory—chiefly, the notion of a stable equilibrium—receive treatment in the following section. After discussing Rawls’s view and clarifying the meaning of “static theory,” I present a challenge for this approach. Section III argues that equilibrium states, also known as fixed points, often won’t exist in complex, dynamic systems, such as society. The standard way of demonstrating the existence of an equilibrium state, Kakutani’s theorem, fails to convincingly demonstrate the existence of justice equilibria. Instead, it points us towards plausible counterexamples in which equilibria do not exist. Static theory therefore faces an existence problem: the focus of analysis may be a phantom, something that never (or rarely) occurs in real social systems. In exploring ways to meet this challenge, part IV suggests that, rather than attempting to prove the existence of an equilibrium state, dynamic justice theorizing should instead focus on nonequilibrium processes. In this approach, the traditional concern with stability is set aside and replaced by the more promising concept of robustness.

6.2 Justice as a Stable Equilibrium

Justice, in Rawls’s view, is “the first virtue of institutions” (1999, 3), that is, the first virtue of our “public system of rules which defines offices and positions with their rights and duties, powers and immunities, and the like” (47-8). A critical desideratum of this virtue is that it be stable. But what, precisely, does stability entail? Stability is a property exhibited by some equilibria. To characterize an equilibrium state, Rawls tells us, requires three things:

... first, to identify the system and to distinguish between internal and external forces; second, to define the states of the system, a state being a certain configuration of its determining characteristics; and third, to specify the laws connecting the states. (1999, 400)

If the successive states of the system, connected by certain laws, are identical, even in the face of external forces acting upon the system, then the equilibrium is stable. Or, as Rawls puts it,

[an] equilibrium is stable whenever departures from it, caused say by external disturbances, call into play forces within the system that tend to bring it back to this equilibrium state, unless of course the outside shocks are too great. By contrast, an equilibrium is unstable when a movement away from it arouses forces within the system that lead to even greater changes. Systems are more or less stable depending upon the strength of the internal forces that are available to return them to equilibrium. (400)

To be an equilibrium—that is, a state \( x \) at time \( t_0 \) that yields the same state \( x \) at time \( t_1 \)—does not suffice to make a state stable. It must satisfy a further property: a state \( x \) is stable if, despite adding some external disturbance \( \epsilon \) to \( x \), at time \( t_0 \) the state \( x \) still emerges at time \( t_1 \). Or equivalently, \((x + \epsilon)\) maps to \( x \). In theory, stability can be quantified: the smaller the difference between \( t_0 \) and \( t_1 \) and the larger the disturbance \( \epsilon \), the greater the stability of state \( x \).
The “system” that Rawls posits as stable is the well-ordered society. The well-ordered society plays a crucial role in Rawls’s work (Kogelmann 2017). It will also feature prominently in the remainder of this paper. So it is worthwhile to carefully review this idea. “[A] society is well-ordered,” Rawls writes... 

...when it is not only designed to advance the good of its members but when it is also effectively regulated by a public conception of justice. That is, it is a society in which (1) everyone accepts and knows that the others accept the same principles of justice, and (2) the basic social institutions generally satisfy and are generally known to satisfy these principles. (1999, 4)²

The well-ordered society presents an ideal of unity and consensus, based on a shared conception of justice. In this society, citizens all share this conception, and the “basic structure,” or society’s fundamental institutional arrangement, is publicly known to be regulated by its principles. In addition, this definition implicitly requires that citizens have a functioning “sense of justice,” a concept that will be discussed later in this section.³

It is tempting to think that stability of the well-ordered society refers to the stability of its basic institutions. The well-ordered society, however, is not defined in terms of any particular institutional arrangement, but rather in terms of its public embodiment of a shared conception of justice. Presumably, this requires that the institutional structure take on a certain form, but in the face of changing circumstances, this need not be the same form in all time periods. Indeed, Rawls explicitly denies that stability of the well-ordered society entails a persistent institutional structure:

>The stability of a conception of justice does not imply that the institutions and practices of the well-ordered society do not alter. In fact, such a society will presumably contain great diversity and adopt different arrangements from time to time. . . . equilibrium and stability are to be defined with respect to the justice of the basic structure and the moral conduct of individuals. (1999, 400-1)

Exogenous shocks—changes in environment, technology, political boundaries, demographics, culture, etc.—may render the basic structure unjust. Following such changes, a basic structure that satisfies the shared conception of justice at time \( t_0 \) may fail to do so at time \( t_1 \). A stable society, according to Rawls, will change the basic structure accordingly to restore its adherence to the conception of justice. The stability that Rawls has in mind, therefore, consists of a potentially infinite array of institutional structures, but a single conception of justice, realized in myriad forms over time as external factors vary. Figure 6.1 depicts this idea of stability.

In figure 6.1, a public conception of justice, presumably justice as fairness in Rawls’s view, gives rise to an institutional structure, \( I_1 \), that embodies this conception given certain background conditions. When these background conditions change, when there is some exogenous shock (dotted arrows), the institutional structure that best embodies a conception of justice is now different; it changes from \( I_1 \) to, say, \( I_2 \). Crucially, each one of these institutional structures promotes, or at least maintains, the

² This is a succinct description of the well-ordered society. The most elaborate definition appears in Rawls’s “Reply to Alexander and Musgrave;” ([1974] 1999, 233-5).

³ Rawls makes this explicit at (1999, 295).
Figure 6.1: The well-ordered society is a stable fixed point, despite constant institutional change.

The well-ordered society is a stable fixed point, despite constant institutional change.

Hence, the well-ordered society is a stable equilibrium in the sense that its underlying public conception of justice remains fixed, despite the perpetual transition from one institutional structure to another.

To take a concrete example, suppose the public conception of justice is Rawls’s preferred conception, justice as fairness. This conception has two key principles: first, the equal assignment of a maximally extensive scheme of basic rights and liberties, and, second, the difference principle limited by equality of opportunity (1999, 53, 266). Suppose that this conception is satisfied at $C_0$ in an institutional structure that allows women full autonomy regarding medical procedures that affect their unborn fetuses. Now suppose that a new technology enables doctors to genetically modify fetuses, raising their IQ, athletic ability, attractiveness, and social intelligence. The institutional structure at $C_0$ will, predictably, give rise to a state where natural endowments are distributed much less equally, undermining equality of opportunity. The institutional structure, in order to maintain its adherence to Rawls’s principles, will need to change. For example, it might regulate the genetic modification of unborn fetuses, it may provide a subsidy enabling less wealthy women to obtain such procedures, or it may provide additional compensatory benefits to those children who did not benefit from genetic modifications. In any case, the institutional structure at $C_1$ will be different from that at $C_0$, even while the conception of justice remains unaltered.

We have now identified two of the three features necessary to characterize a stable equilibrium state. The system under consideration is the well-ordered society. The relevant states of this system are not its particular institutional forms, but rather the personal and institutional adherence to a public conception of justice. The final feature of this characterization, the laws that link the states, are presented by Rawls in the form of an elaborate moral psychology. The centerpiece of Rawls’s psychological theory is one crucial concept, mentioned above: the sense of justice.

The sense of justice provides the main mechanism by which justice at time $t_0$ gives rise to justice at time $t_1$. It is this concept that explains why
individuals who live under just institutions will strive to maintain the justice of their institutions and to reform these institutions when they deviate from justice, whether this deviation be due to institutional decay or due to changing circumstances (1999, 415). According to Rawls, various psychological laws account for the emergence of a sense of justice, and he argues that his favored conception of justice, justice as fairness, fosters a particularly strong sense of justice in light of these laws. Examining this moral psychology in detail would lead us far afield. What matters is this: the sense of justice ensures stability by motivating citizens in the well-ordered society to devote effort and resources to maintaining the justice of their institutions, as specified by a public conception. Hence, the sense of justice, and the psychological laws that ungird it, provide the third crucial component of characterizing an equilibrium: the laws connecting the states.\(^5\)

In sum, we have identified the three general components of Rawls’s justice equilibrium:

1. The System: The well-ordered society.
2. The States: The prevailing public conception of justice, which I refer to as a “justice state.”
3. The Laws: The sense of justice and its generation from institutions via the three psychological laws.

Put differently, Rawls posits that social and psychological laws provide a function, call it \(f\), that maps from the set of justice states into that same set. Moreover, he posits that, given the laws that this function expresses, the well-ordered society presents a “fixed point,” i.e. a state within the set that maps to itself via \(f\). In mathematical notation, which will be helpful in the next section, if the point \(x^*\) represents the well-ordered society, then \(f(x^*) = x^*\). This idea of a fixed point is simply a way of formalizing the idea of an equilibrium.\(^6\)

Rawls does not merely assert that the well-ordered society is a fixed point. Equilibria can be stable or unstable. Hence, Rawls asserts the additional claim that the well-ordered society is a stable fixed point. The ensuing discussion of Rawls’s work has focused on the concept of stability. The present argument, however, focuses on the prior concept of equilibrium. Before engaging in an analysis of the properties of justice equilibria, the political theorist owes us some account of why we should expect justice equilibria to exist. It is to this neglected question that we now turn.

### 6.3 The Problem of Existence

Karl Popper influentially argued that, as a complex open system, society will not equilibrate. “Society,” he claims, “is changing, developing. Its development is not, in the main, a repetitive one... Conditions are changing, and situations arise (for example, in consequence of new scientific discoveries) which are very different from anything that ever happened before” (Popper 1963: 457). Many historians share Popper’s belief, eschewing the search for general laws of history, and, as we will see, some have even extended this stance to conceptions of justice. How do we know that any conception of justice, including justice as fairness, can support an equilibrium in the form of a well-ordered society? In

---

4: There are three psychological laws that Rawls highlights. For a detailed account of the first law and its connection to the family, see (1999, 406). For a detailed account of the second, see (411-2), And for the third, see (415). For a concise summary of all three, see (430) and for their connection to the sense of justice, see (436).

5: Importantly, the sense of justice is not the only mechanism that allows the well-ordered society to generate its own support (Rawls 1999, 211). Nevertheless, Rawls’s account of stability focuses predominantly on the sense of justice, as do most of his commentators (Weithman 2010; Thrasher and Vallier 2018). I follow suit.

6: Rawls suggests that such formalization is legitimate: “these ideas admit of considerable theoretical and mathematical refinement but I shall use them in an intuitive way” (400). And as Wolff (1977) points out, contemporary developments in mathematical economics heavily influenced Rawls’s theory.
other words, what does it take for an equilibrium to exist, and under what conditions might we suspect that it does not?

The standard approach in the social sciences is to apply Kakutani’s Theorem. From game theory to general equilibrium theory, social scientists rely on Kakutani’s Theorem when demonstrating the existence of equilibrium states in a model system. This theorem specifies a set of conditions, which, when satisfied, guarantee that there exists at least one fixed point as a mathematical fact. Understanding how these conditions might fail to be satisfied will be instructive in assessing the plausibility of the static theorist’s assumption that there exists some fixed point of justice.

For our purposes, we can provide an interpretation of Kakutani’s Theorem that translates it into a theorem about justice:

- Let $A$ be a set of points in $\mathbb{R}_n$ that contains all possible “justice states” of society, with $n$ being the number of justice-relevant parameters. Hence, every state of society is described by a vector of length $n$, which is why $A$ is a subset of $\mathbb{R}_n$.
- $f$ is a function that maps one justice state, any point in $A$, to another point contained within $A$.
- There are four conditions required for Kakutani’s Theorem to apply:

  (C1) Suppose the set of possible justice states, $A$, is compact: it contains its own boundaries. That is, the extreme values that any of the $n$ dimensions might take can be realized ($A$ is closed), and these $n$ dimensions do not extend outward to infinity ($A$ is bounded).
  
  (C2) Suppose that the set of possible justice states $A$ is convex: it includes all intermediate states between any two of its possible states.
  
  (C3) Suppose $f$ has a closed graph: if $f$ can map arbitrarily close to some justice state $x'$, then there is some justice state $x$ that will map directly onto $x'$.
  
  (C4) Suppose $f(x)$ is non-empty for all $x$ in $A$. That is, a justice state $x$ always maps to one (or more) other justice states via $f$, even if this means that the state $x$ maps to itself.

Given our interpretation of Rawls’s theory (sec. II), a “justice state,” $x$ refers to some publicly held conception of justice. The transition function $f$ refers to the process that takes society from its public conception, through the institutional manifestation of that public conception, and back to a new public conception of justice. When (C1) - (C4) are satisfied, a fixed point must exist: there will be some well-ordered society, under some public conception of justice, that is in equilibrium, meaning that it maps onto itself.

Graphically, a fixed point will be any point that falls along the 45 degree line represented by $g(x) = x$ (the dashed diagonal line in figure 6.2). If a function, representing a system, can be drawn that relates all original states to new states without intersecting the dashed line $g(x) = x$, then the system has no fixed points. One way of understanding Kakutani’s Theorem is as a set of conditions which ensure that $f(x)$ will, in at least one place, intersect the line $g(x) = x$. 


Is Justice a Fixed Point

Note that, while these conditions are sufficient for a fixed point to exist, they are by no means necessary. However, relaxing any one of these conditions allows for counterexamples: we can find some system in which no equilibrium exists. If these counterexamples, once interpreted in concrete terms, sound like plausible descriptions of our own social-moral systems, then we have reason to doubt that there exist any fixed points of justice. The argument, therefore, aims to shift the burden of proof. If there are plausible construals of society in which no justice equilibrium exists, then the theorist may not simply assume that these fixed points of justice exist. The theorist must prove this, or at least provide some reason for us to believe it.

Let’s start with the simplest possible graphical representation of a fixed point by assuming that \(n = 1\). That is, there is only one dimension that is relevant to justice. Moreover, we can measure this justice-relevant feature on a scale of 0 to 1. Thus, let \(A = [0, 1]\), and let \(f : [0, 1] \rightarrow [0, 1]\) be defined so that \(f(x) = 2x^2\). It is easy to check that this system, represented in figure 6.2, will satisfy all of the conditions of Kakutani’s Theorem.

One possible interpretation takes \(x\) to be the level of some dimension of justice, say, social equality. In this case, 0 represents the limit of extreme inequality, and 1 represents full equality. On the \(x\)-axis are arrayed all possible levels of social equality that a society might endorse as its public conception of justice. The \(y\)-axis represents all of the new states that the original states might map onto. The solid line represents the laws of transition, that is, the function \(f\) shows how an original state will map onto a new state. Again, the dashed line, \(g(x) = x\), is simply there for reference: any point falling on the dashed line will, by definition, be a fixed point, since on this line the original state and the new state are equivalent. Our particular function \(f\) says that when society is unequal, i.e. \(x < .5\), social equality tends to decrease until society is entirely despotic. However, \(f\) also shows that equality is self-reinforcing in the sense that a society with a high measure of equality will tend to increase in social equality. From Kakutani’s Theorem, we know that our graph
must possess at least one fixed point. Intuitively, this fixed point should occur when the forces pulling us towards less social equality perfectly balance with those pushing us towards greater social equality. As we can see, this fixed point occurs at \( x = 1/2 \). In fact, there exist two fixed points, since the equation: \( f(x) = 2x^2 = x \) has two solutions: \( x = 0 \) and \( x = 1/2 \). At both of these points, \( f \) maps \( x \) onto \( x \). The well-ordered society is in equilibrium when there is either an intermediate level of social equality or when society is entirely unequal.

This example is helpful for visualizing a fixed point, but it is rather unrealistic. What happens, for example, above \( x = 1/2 \)? According to the graph, equality will continue to increase indefinitely, since the “new state” is above the dashed line, i.e., the new state is more equal than the original state for all \( x > 1/2 \). But then, the original state \( x = 1 \) will map to some new justice state, \( x \), that’s greater than 1. Given that equality is measured on a scale of \([0, 1]\), this model hardly makes sense. Let’s consider another example. For instance, suppose \( f(x) = x^2 \) instead of \( 2x^2 \). Then we get the following:

In this new system, the only two fixed points are total equality or total despotism. Now we can begin to cash in on all of this formalization. Conditions (C1) and (C2) concern the nature of the set of possible justice states, that is, the shape and properties of the set of public conceptions that a society may come to endorse. Conditions (C3) and (C4), on the other hand, concern the nature of the transition function \( f \), that maps from one justice state to another. Rather than considering a counterexample for each condition, we will examine three potential cases, one where the set of justice states fails to satisfy the conditions of Kakutani’s theorem and two where the transition function fails to do so.

Consider what happens if our society fails to satisfy (C1). Recall that (C1) stipulates that our space of possible justice states, previously \([0,1] \), must be compact, which means, roughly speaking, that it contains its (finite) boundaries. To violate (C1), let’s replace \([0,1]\) with the very similar, though
Is Justice a Fixed Point

Figure 6.4: By violating (C1), we have eliminated the fixed points at (0, 0) and at (1, 1).

not compact, space (0,1). In terms of our example, this is tantamount to supposing that there is no way for society to occupy a state of equality with a score of 0, i.e. total despotism, or a score of 1, i.e. perfect equality. We can get extremely close — arbitrarily close even — but, due to the confluence of myriad social and natural facts, the universe simply will not permit a justice state to realize perfect (in)equality. The result is figure 6.4:

Notice that both of our fixed points have vanished: the points (0,0) and (1,1) have been replaced with empty circles to show that these are no longer valid inputs or outputs. With an extremely minor change in the nature of the set of possible justice states — we reduced this set by an immeasurably small amount — we have eliminated the existence of any fixed point of justice. Moreover, the supposition employed to do so is not unrealistic. It’s not difficult to imagine that we can approach full equality or full despotism without ever completely realizing either.

Perhaps the counterexample depicted in figure 6.4 seems to be of mostly theoretical interest, since the system will asymptotically approach the point (0,0), even if it never arrives. This seems, on the one hand, unrealistic, and on the other, pretty similar to a fixed point. But this is a weak interpretation. Perhaps the story accompanying the graph is that, however close we get to the ideal of total equality, society is constantly pulled closer and closer to total despotism, and dissolves upon reaching this state. Under this plausible interpretation, the dynamics depicted in figure 6.4 accord with highly influential political theories that posit a degeneration into tyranny. For instance, one might read Book VIII of Plato’s Republic in just this manner. Another example is Marx’s theory that the capitalist class will control an ever greater share of wealth, while imposing a capitalist ideology on the populace, ultimately resulting in a revolutionary regime change. In these theories, society is drawn towards tyranny, but total tyranny is not a fixed point, since society will not subsist (or will be radically altered) upon reaching this state.

8: Technically, society will never reach 0, but it will get so close that even infinitesimally small error or noise will push it there.
A very similar counterexample serves to demonstrate that violating (C2) also permits systems lacking fixed points of justice. However, different and more interesting counterexamples arise when we focus on conditions (C3) and (C4), which concern the transition rule \( f \). If this transition rule is undefined for certain states or discontinuous at certain points, a system may not exhibit any fixed points. What would these abstract properties look like in a real social-moral system? Gustav Schmoller offers one possible interpretation:

...the same theory which proposes a demand of justice as its consequence often is made by individuals, but repudiated by public opinion; and then suddenly with irresistible elementary force it takes hold of the masses, leads them on new paths, radically influences legislation and puts a changed stamp on whole epochs. (Schmoller 1894, 712)

A public conception of justice might change slowly and continuously, like a stick gradually bending into an arc, until it reaches a critical point where the stick snaps, disrupting the continuous trend that preceded this new state. The snapping of the stick may correspond to the dissolution of anything recognizable as a “public conception of justice,” or it may correspond to an entirely new trajectory along which the public conception of justice will evolve. There is no way to rule out the possibility of such social-moral dynamics a priori, and indeed, a cursory glance at social movements reveals fits and starts that belie the assumption of smooth continuity.9

This idea of building pressure and sudden change corresponds to a common violation of Kakutani’s theorem, namely (C3), which demands that the graph of \( f \) be closed, i.e., it should contain all of its endpoints. In particular, one way to violate this condition is to introduce a discontinuity in society’s conception of justice. Perhaps, to take a simplistic example, society is committed to equality even if equality is declining for some lengthy period of time. There may come a critical point, however, call it the “threshold point,” when society undergoes a major transition, and finds itself pulled toward a completely different basin of attraction.10

In figure 6.5, there are no fixed points. A discontinuity occurs at \( x = 0.5 \), where society finds its public conception of justice to be radically transformed. Supposing the .5 threshold is crossed, a different mindset takes over and citizens become more prone to rejecting equality. At no point does the system equilibrate. There is no justice state that maps onto itself.

The introduction of discontinuity provides a graphical representation of Schmoller’s description of sudden shifts in public conceptions of justice. Again, our one-dimensional analysis of justice is highly stylized and rules out more interesting dynamics. As depicted, a society approaching \( x = 0.5 \) from the right would leap to a new state, located on the other curve, and then forever flutter about without settling down. Although more interesting dynamics would be possible with a higher-dimensional model, the representation allows us to visualize one of the ways in which a public conception might fail to equilibrate.11 Systems with discontinuous dynamics may lack equilibria.

9: Many historical examples of discontinuous transitions are examined by Kalvyas (2006).

10: Aligica and Tarko (2014) discuss “slippery slopes” that lead to discontinuities, or “catastrophic thresholds.”

11: A multidimensional model is developed below.
Schmoller’s view that society exhibits discontinuous changes finds support in contemporary social science. Following Schumpeterian lines of reasoning, complexity economists have argued that endogenous novelty prevents equilibrium states from arising in the economy. “[N]onequilibrium,” Brain Arthur assures us, “is the natural state of the economy” (Arthur 2015, 5). Just as the economy might be approaching a fixed point, a new technology, product, or organizational form appears, and “a gale of creative destruction” radically alters the economic situation (Schumpeter 2008, 84). Analogously, we might imagine cultural or “ideological entrepreneurs” who produce “discontinuous change” in our public conception of justice (Storr 2009, 105).

There is one more counterexample worth considering. Focusing on (C3), we might relax the assumption that, for any justice state \( x \) in the domain \([0,1]\), \( f(x) \) is defined. That is, some justice states will fail to map onto another justice state. Violating this condition might be interpreted in many ways. Some states may lead to the disappearance of any and all consensus on justice, so that a society in such a state simply ceases to possess any recognizable public conception of justice. Or perhaps a conception of justice is so corrosive that its wide adoption among the populace leads to the dissolution of society altogether. Alternatively, remaining within a Rawlsian framework, an undefined area in the domain could indicate that one of the subjective or objective circumstances of justice fails to hold (Rawls 1999: 109-10). According to Rawls, a conception of justice would then have no role to play, and citizens would not require (or be able to instantiate) principles of justice in their institutions. There are other possible interpretations, as well, but this suffices to show that condition (C4) is not guaranteed to hold for all social systems, even within Rawls’s conceptual schema.

The last two counterexamples were confined to a one-dimensional justice space, greatly limiting the behavior of the system. Adding more dimensions would allow for a much larger variety of non-equilibrium behaviors. In a scenario with two justice-relevant dimensions, for ex-
Is Justice a Fixed Point?

ample, a violation of (C4) could involve a scenario in which the justice states continually oscillate around some point, never settling down into a static equilibrium state. Although it will unfortunately complicate the simple model, a representation of the two-dimensional scenario will also provide substantial insight.

An illuminating graph of the new dynamical system calls for a different kind of plane, known as the “phase space.” Rather than placing the current state on the x-axis and the new state on the y-axis, as above, the new plane places one justice-relevant parameter on the x-axis and a different justice-relevant parameter on the y-axis. Lines on the graph, accordingly, represent trajectories of two-dimensional states as they evolve over time. The point of this new plane is to represent the relationship between the two variables. For this illustration, let us select two justice-relevant variables to which Rawls grants paramount importance. Let F represent the first principle of justice and E represent the second principle of justice. For purposes of exposition, let F be referred to as freedom and E as equality. To keep the equations simple, we will measure both F and E on a plane centered at the origin. Suppose, as is likely the case, that the extent to which freedom and equality shape society’s institutions affects how citizens view and prioritize these two values. The simplest relationship of this kind would result in a linear system of differential equations:

\[
\begin{align*}
\frac{dE}{dt} &= aE + bF \\
\frac{dF}{dt} &= cE + dF
\end{align*}
\]

where \(a, b, c, d\) are real parameters, and the left hand sides of each equation represent the change over time of equality (E) and freedom (F), respectively. In this scenario, the transition function \(f\) will take two inputs, one value for E and one for F, and will map to a new point in the next period. The change induced by \(f\) will no longer be represented by the difference between the \(x\) and \(y\) values, but by arrows indicating the direction of a trajectory. For different parameter specifications, the system (\(\ast\)) will behave in very different ways. For instance, if we fix \(a = d = 0\), and if \(c\) and \(b\) have opposite signs, then the system is guaranteed to exhibit cycles. The origin will always be a fixed point, but given that we have relaxed assumption (C4), \(f\) need not be defined for all input pairs. Consequently, if \(f\) is undefined for any region containing the origin, this system will not exhibit any fixed points, only cycles. There are, therefore, infinitely many possible systems exhibiting cycles and lacking equilibria (figure 6.6).

In broad outline, figure 6.6 conveys a story in which a greater preference for freedom produces a greater preference for equality. However, a greater preference for equality diminishes citizens’ preference for freedom. Recalling the Rawlsian framework, the details of the story depicted in figure 6.6 will involve a bidirectional interaction between the public conception of justice and its institutional manifestation. Perhaps freedom is instantiated as economic liberalization, which creates socio-economic inequality, thus producing a reaction in thought and feeling that causes a greater desire and higher esteem for egalitarian values. And perhaps
equality is instantiated in a more communitarian social organization that suppresses the desire for individual liberty. The exact details need not occupy us here. Instead, the key point is that the possibility of cyclical systems lacking equilibria calls upon static theorists to provide a more direct defense of the assumption that social-moral systems will contain fixed points.

Even this more complex model is, of course, highly simplified. Rawls’s theory, as well as other static theories, include many intricate features not represented here. For example, both Rawls’s liberty and equality principles are multi-part and are regulated by certain priority rules (1999: 266). Complications — for example additional variables or non-autonomous terms — could be introduced for the sake of realism or for textual fidelity. Most promisingly, the introduction of nonlinear terms would permit the emergence of more complex behavior, such as limit cycles, which aptly model a wide variety of cyclical phenomena (Strogatz 2015: 198). Though worthy of closer study, these more realistic models are unnecessary for the present argument. The point of this formalization is not to definitively prove that Rawls’s conception of justice — or any other for that matter — will never equilibrate. Instead, the point is to indicate that static theorists need to justify the assumption of a fixed point. There are many scenarios in which a fixed point will not exist, some of which involve perpetual cycles.

Although this model is highly stylized, these stylized scenarios are analogous to processes described more richly by political theorists and social scientists. Political theorists should not dismiss the cyclical view of society as empty formalism, for such views hold a place of prominence in the history of political thought.13 In fact, such cyclical oscillations provide a dynamic representation of one of the most influential results in political theory: Arrow’s impossibility theorem (Arrow 1970). If a society continually cycles between several justice states, then the social choice function necessarily violates transitivity. A more intricate version of the model in figure 6.6 could thus depict the scenario in which society, preserving Arrow’s four conditions, violates transitivity in determining its public conception of justice. As many have noted, the prospect of

---

intransitive social choice raises interesting normative questions. Kakutani’s Theorem has guided us in imagining simple scenarios where the sort of system envisioned by Rawls will fail to equilibrate. As argued above, one cannot begin to discuss the stability of an equilibrium until one has established that the equilibrium actually exists. Static theorists of justice thus face a challenge: they must justify the assumption of equilibrium. The most common way of demonstrating the existence of an equilibrium, viz. invoking Kakutani’s Theorem, leads to difficulties. As this section has argued, the conditions of Kakutani’s Theorem plausibly fail to hold in many scenarios, allowing for perpetual disequilibrium in justice systems. Perhaps there is some way to meet this challenge and, thus, to justify static theory. However, the next section argues that there is a promising alternative. Examining justice as a non-equilibrium process opens up a new avenue of research with the potential to yield novel insights into the nature of justice.

6.4 Dynamic Theorizing

In contrast to views that examine the stability of different conceptions of justice, the present paper has asked a different question: will the well-ordered society, if realized, be in equilibrium? In many situations, there will be no fixed point of justice. Society will flout the conditions of Kakutani’s Theorem, opening up the possibility that no public conception of justice can be iteratively reaffirmed, even in the absence of exogenous shocks. Social systems, as many theorists have argued, may exhibit constant, endogenous disequilibrium. As John Miller and Scott Page put it, “Stock markets soar and crash… Political parties rise and topple…” — and conceptions of justice, I would add, surge and recede — “While the notion of social equilibria is an important one… we may need to go beyond equilibria to truly understand the social world” (Miller and Page 2007, 222).

Given the prevalence of endogenous disequilibrium in social systems, how do we theorize about justice? For those committed to focusing exclusively on static justice equilibria, there are two possible responses. One approach is to drastically narrow the focus of political theory. By restricting attention to cases in which justice equilibria actually exist, and can be shown to exist, justice theorizing may proceed as normal, with the qualification that it focuses only on these special cases. If continual autocatalytic change characterizes contemporary societies, however, then these cases will be the exception rather than the rule (Beinhocker 2006: 70-4, 99). The major drawback of this approach is, first, the difficulty of demonstrating the existence of equilibria, and second the highly contextual nature of any conclusions or prescriptions that the theorist offers. In being so constrained, static models will lack what scientists call “external validity.” That is, even if a static theory identifies the correct principles of justice for a particular time or place, changes in the justice-relevant variables imply that such theories will not generalize to different contexts.

Another possible response is to impose conditions on society that make it more amenable to static theorizing. This may be what Rawls had in mind
when he entertains the notion of a steady state economy ([1993] 2001: 107, n.33). Commentators on Rawls have noticed the inconsistency between the notion of stable equilibrium and the dynamism of our contemporary, globalized economy (Schrepfer 2019: 159). Thus, one possible response to the endogenous change exhibited by today’s complex, globalized social systems is to suppress social complexity, as many states historically have (Scott 1999; Hall 1985). By reducing the diversity and interdependence of contemporary societies, we might foster the existence and stability of static justice equilibria. The major drawback of this approach is that such simplifications will likely come at severe cost, not just to our material well-being, but also to our political freedom. Indeed, later Rawls concedes that a diversity of perspectives spontaneously arises as the natural result of free institutions and that the suppression of this process could come about only through the “oppressive use of state power” (2005: 37). The cost of facilitating stable fixed points, therefore, may be prohibitive.

Assuming that broad applicability is a desirable feature of a political theory and that the radical simplification of contemporary society is undesirable (or unattainable), then we must ask if there is an alternative approach to theorizing that adequately responds to the problem of existence. Indeed, there is: we may explore methods of theorizing that do not assume the existence of equilibria. The approach considered and criticized in section III, which I have called static theory, studies conceptions of justice insofar as they exhibit equilibrating properties. By contrast, the alternative approach, dynamic theory, will examine social-moral systems exhibiting continual flux. These two approaches differ in their objects of study and, consequently, in their preferred methods.

With respect to the objects of study, static theories of justice consider the array of social values to be either exogenous or self-reinforcing and, thus, changeless. These values are held by citizens who must formulate just terms of engagement to facilitate cooperation. Rawls’s version of static theory strives for a “reflective equilibrium,” in which the principles underlying society’s basic institutions match our coordinated, considered judgments (Rawls 1999: 17-8). The static theorist thus focuses on a justice state, seeking to understand the institutions that will best instantiate and support the hypothesized justice state. By contrast, dynamic theory accepts that values are endogenous and evolving. The dynamic theorist thus focuses on how these constantly changing values can be accommodated and adjusted to maintain cooperation over time. That is, the dynamic theorist focuses on a justice process, seeking to understand how institutions will accommodate coordination amidst a network of agents whose considered judgments are reflexively adapting to the political and social structures they help to produce. Dynamic theory examines systems that permit constantly evolving rules in response to constantly evolving values, thus eschewing the assumption of fixed point equilibria of justice.

Given these different objects of study, the two types of theory employ very different methods. Static theorists tend to be utopian, because they take their task to involve characterizing a stable social equilibrium that instantiates the given values of its members. Again, taking Rawls as the exemplar, the assumption of full compliance is introduced, since a fully just society will reinforce its own underlying values, constituting a fixed point of justice. The project that Rawls undertakes is thus “realistically

---

14: See, also, Rawls’s discussion of “general” reflective equilibrium in Political Liberalism (2006: 384, fn.16), which emphasizes the interpersonal, coordinated aspect of reflective equilibrium.
Is Justice a Fixed Point?

15: In some construals of the debate, this means that while static theorists skew ideal in their methodology, dynamic theorists skew nonideal. See Valentini (2012).

Furthermore, since stability as a concept presupposes equilibrium, it is unavailable to the dynamic theorist. Rather than stability, therefore, the dynamic theorist should focus on the analogous concept of robustness. Thrasher and Vallier (2018) have recently offered an important exploration of this concept. For Thrasher and Vallier, robustness is not a feature of a system in equilibrium, but of a process of equilibration: a robust system “tends toward equilibria” (406). This is a major advance over prevailing static theories, but the reference to equilibria is insufficiently radical, since, as argued above (sec. III), social systems may not exhibit any equilibrating tendencies. Luckily, referencing equilibria is neither necessary nor standard in definitions of robustness. In general systems theory, robustness often refers to a system’s ability to “maintain functionality” in the face of shocks (Page 2010: 8). Put differently, system robustness refers to “feature persistence” (Jen 2005: 10). A system is robust, in other words, when, despite exogenous and endogenous shocks, it maintains some of its qualitative features, even as it shifts between equilibria or remains in a non-equilibrium state. Accordingly, the task of the dynamic justice theorist is to identify desirable features that a process of justice might possess and to determine whether these features will be robust as society shifts between distinct states of justice.

As an example, consider the multivariable model depicted in figure 6.6. Here, society exhibits a constantly evolving conception of justice, one which goes through periodic cycles. A dynamic theorist, without assuming any fixed point conception of justice, may consider the flexibility and revisionary nature of this social-moral system to be a normatively desirable property. In fact, in the context of discussing the intransitivity of democratic procedures, James Buchanan defended the normative appeal of cyclical social choice. By allowing diverse perspectives to prevail at different points in time, the formal inconsistency of such collective decisions “provides one of the most important safeguards against abuse” (Buchanan [1954] 1999: 96). Moreover, policies induce changes in beliefs and preferences, and “[i]ndividual values are... constantly changing; so a postdecision ordering may be different from a predecision ordering” (Buchanan [1954] 1999: 99). If, following Buchanan, the dynamic theorist holds that cycles, or intransitivity in general, is a desirable feature of social choice, then the next step is to ask whether or not the social-moral system is robust: will this desirable property persist as the public conception of justice evolves? For the system depicted in figure 6.6, the answer is “yes.” Although an exogenous shock will shift the system to a new trajectory, it will remain cyclical, thus preserving the intransitive nature of social choice. If the answer were “no,” the dynamic theorist would consider which institutional reforms could secure greater robustness.

This example reinforces the point expressed by Holling in the epigraph about the advantage of a dynamic approach. Whichever (defined) point in the phase space the theorist selects as most just, society will resist any utopian,” since it aims to determine how “a democratic regime can attain complete realization of its appropriate political values” (2001: 13). For the dynamic theorist, there is no utopian social order. Instead, the basic fact of politics is disagreement and disputation: as values evolve and change over time, even a (previously) ideal social order will prove to be inadequate. Rather than a state without errors, dynamic theorists seek a process of error correction.15

Is Justice a Fixed Point

For instance, about half of the ten commandments seem to be cross-cultural. Attempt to pin it there. To the static theorist, therefore, the fluctuating and inconsistent system could only be viewed as a messy failure to instantiate or maintain a state of justice. By contrast, the dynamic theorist observes features of the process itself. This adds an additional dimension to the analysis; what appears irrational or immoral to the static theorist manifests an underlying order and value to the dynamic theorist. The static theorist sees only the failure to remain at the correct state, while the dynamic theorist sees a process of error correction in response to evolving values.

Although static theory, on its own, is deficient, this does not mean that there is no value to static theory or that the static theorist and the dynamic theorist are necessarily antagonists. To the contrary, these two approaches can complement one another. The important distinction between processes and states parallels a distinction between process desiderata and state desiderata. Some, though not all, state desiderata are sectarian in that they are internal to a particular conception of justice. For example, in a strict egalitarian conception, a society is more just to the extent that it prohibits unequal material holdings. On the other hand, some state desiderata are general to all conceptions. For example, reducing violent conflict is a desirable feature of a conception of justice whether one is an egalitarian, a libertarian, a conservative, a socialist, or whatever. The static theorist can complement the dynamic theorist by identifying such general desiderata, and determining the institutions that enable them to persist as stable equilibria. While it is unlikely that whole conceptions of justice will constitute stable equilibria, certain values seem to transcend time and place, even in cases when our institutions sadly fail to instantiate these values.

Sometimes these transcendent values are process desiderata, in which case they fall within the purview of the dynamic theorist. Sometimes, on the other hand, these transcendent values are state desiderata, in which case the static theorist can make an important contribution. On this analysis, static theory can make an important contribution, but historically it has made a twofold error: (1) an error of commission in thinking that whole conceptions of justice, rather than a few general values, will likely remain in equilibrium, as well as (2) an error of omission in neglecting the nonequilibrium processes that, I have argued, are crucial to understanding justice.

Although I have cast Rawls as the archenemy of the dynamic approach outlined here, this is true only up to a point — in particular, up to 1993. Rawls ultimately came to reject part III of A Theory of Justice, because “the account of stability in part III of Theory is not consistent with the view as a whole” (1996: xvii-xviii). The citizenry of a liberal society, Rawls came to believe, is too diverse to embrace a single conception of justice as a “comprehensive” philosophical doctrine (xxxvii, 13). Ultimately, Rawls shifts the focus away from his preferred conception of justice, viz. justice as fairness, instead emphasizing the process by which we come to embrace a shared set of social-moral rules. That is, he shifted his focus from the outcome of a social process of coordination to the evaluation of the process itself. Society is just when this process conforms to the meta-rules of public reason, a set of reasonable constraints on discourse that allow widely diverse individuals to engage in a shared process of interpersonal justification despite their deep moral disagreements. In

17. For instance, about half of the ten commandments seem to be cross-cultural.
this regard, by the end of his career, Rawls’s theory was taking on an increasingly dynamic character (Gaus and van Schoelandt 2017; Gaus 2013). Post-Rawlsian political philosophy has seen a burgeoning interest in dynamic theorizing about justice. A central figure within this new program, the late Gerald Gaus, explores the implications of complexity, diversity, and dynamism in several works, most notably The Tyranny of the Ideal (2017). Gaus argues that in the process of pursuing an ideal, we learn more about our institutions and our values. As a result, individuals must reevaluate the features of that ideal, as well as their criteria of evaluation. What was once an ideal no longer is; the “fitness landscape” has shifted, and the search must go on.

According to Fred D’Agostino, Gaus, along with several others, forms part of a “new program,” which “more or less abandons the project of providing an end-state description of a justified social order” (D’Agostino 2016: 31). Instead, these theorists focus on processes that enable stakeholders to instantiate, legitimate, and continuously reevaluate their own social and political arrangements. Future work on dynamic theories of justice promises to be one of the most fruitful and exciting developments in post-Rawlsian political thought.

6.5 Conclusion

Charles Lindblom is known partly for what he calls the “branch method” of policy choice. At the foundation of his defense of this method is the claim that, in most complex policy decisions, means and ends are intertwined; values and policies are chosen simultaneously. This is true in at least one sense that is important for justice theorizing: a political conception of justice and our institutional framework exhibit mutual causation. A citizen’s view of public morality will affect her support or rejection of constitutional and institutional rules. Similarly, these constitutional and institutional rules will affect the material conditions of a citizen’s life, including her social and political experiences, thus affecting her evaluation of public-moral rules. The consequence, according to Lindblom, is that our criteria of evaluation evolve alongside our policies and institutions.

This paper has adopted this basic insight to argue that static justice theorizing faces severe challenges in a dynamic system characterized by constant disequilibrium. Theorists such as (early) Rawls have insisted that the requirements of justice are static and unchanging, even if their institutional manifestations may shift over time. However, drawing on Kakutani’s Theorem, it is easy to imagine cases where the equilibrium, far from being stable, will not even exist. How can the justice theorist accommodate the possibility of nonequilibrium?

I have argued that the key is to focus on robustness, rather than stability, where robustness is understood as maintaining certain process desiderata in the face of changing institutions and conceptions of justice. This requires the justice theorist to identify these desiderata and to study diverse institutions and their respective abilities to satisfy these desiderata. Doing so will allow her to illuminate an otherwise obscure aspect of any

---

18: Especially notable here are Ryan Muldoon (2016) and Julian Müller (2019), who have both contributed book-length explorations of dynamic theory.

19: In other words, the social system comprising these levels is not decomposable (D’Agostino 2009).

20: As Rawls puts it, “the basic institutions of society . . . can have decisive long-term social effects and importantly shape the character and aims of the members of society, the kinds of persons they are and want to be” (2001, 326). See also (Rawls 1996, lec. VII).
system of justice. In particular, the process of disputation, adjudication, and resolution comes to the fore.

As a final consideration favoring the dynamic approach to theorizing, consider the importance of this process in our intuitive evaluations. What we typically regard as the most heinously unjust societies are not the ones that fail, in some given “snapshot,” to satisfy a criteria of justice, or to successfully adjust the beliefs and values of their citizens at some given period in time (Schmidtz 2008, 6-7). Heinous injustice occurs when the process of adjustment, i.e. the process of justice, ceases to function. If groups can air their grievances and communicate their expectations, there will be a tendency toward appeasement, compromise, and reconciliation. When groups are marginalized, disenfranchised, censored, or silenced, the tendency is, instead, toward exploitation and oppression. What saves society from such heinous injustice isn’t the snapshot equilibrium, it’s the existence of a process by which individuals and groups can express frustration and make amends. Often, it is this process, not the snapshot, that we call “justice.”
Appendix

A.1 Deriving Price’s Equation

Set-up

- There is a population $P$ consisting of $n$ entities.
  - ‘$P$’ stands for ‘Parent’.
- $z$: A (phenotypic) trait of some kind, which can be measured (discretely or continuously) with real numbers.
- $z_i$: The level of $z$ exhibited by the entity $i \in P = \{1, \ldots, n\}$.
- $\bar{z}$: The average level of $z$ in the $P$-population.
  - Mathematically, $\bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$
- $z_i'$: The amount of trait $z$ transmitted by $i$ to any of its “offspring.”
- $\Delta z_i$: The change in the level of $z$ from one entity to its offspring.
  - $\Delta z_i = z_i' - z_i$
  - $z_i'$ can be thought of as the copying fidelity or transmission bias of trait $z$ for entity $i$.
  - Taking an average across all entities in $P$, the expectation of $\Delta z_i$ is $E[\Delta z] = \frac{1}{n} \sum_{i=1}^{n} \Delta z_i$
- $w_i$: The average number of offspring produced by entity $i$.
  - $w_i$ can be thought of as the “fitness” of entity $i$.
  - Similarly to $\bar{z}$, we define the average fitness as $\bar{w} = \frac{1}{n} \sum_{i=1}^{n} w_i$
  - Often we will be interested, not in the absolute fitness of an entity, but in its fitness relative to the other entities in the group. For this comparative purpose, we define relative fitness of entity $i$ as $\omega_i = \frac{w_i}{\bar{w}}$.
- There is another population of interest, that comprised of the offspring of all entities in $P$. Call this population $O$.
  - ‘$O$’ stands for ‘Offspring’.
- $\bar{z}_o$: The average level of the $z$ trait in population $O$.
  - Mathematically,
    - $\bar{z}_o = \frac{1}{n} \sum_{i=1}^{n} \omega_i z_i' = \frac{1}{n} \sum_{i=1}^{n} \frac{w_i}{\bar{w}} z_i'$
    (A.1)
  - Importantly, for the derivation, $E[\omega] = \frac{1}{n} \sum_{i=1}^{n} \omega_i = 1$, which makes sense given that $\omega_i$ is a proportion of the total $O$-population.
The Partition Theorem

Equation A.1, while it describes the O-population, is couched entirely in terms of P-population traits. This may seem somewhat mysterious, but it is the most crucial equation to understand for the derivation that follows. This equation is easy to grasp once one understands the so-called partition theorem. The idea is actually quite intuitive. Suppose we divide the total population into a set of jointly exhaustive and mutually exclusive subpopulations. Then the average level of a trait, e.g. $z$, within the population as a whole will simply be the (weighted) sum of the averages of each subpopulation, where each one is weighted by its relative size. This idea can be illustrated with a simple diagram.

In figure A.1, we have divided a population comprised of dots into four subpopulations, $A_1 - A_4$. We might think of each of these subpopulations as the offspring of a single entity in the P-population, so that $A_1$ consists entirely of entity 1’s offspring, $A_2$ of entity 2’s, and so on. Suppose the size of each dot represents the magnitude of our quantity of interest (here, the amount of trait $z$ each entity in the O-population has). The partition theorem states that we should add up the average $z$ level in each of the subpopulations $A_1 - A_4$, weighting each of these by the proportion of the total entities that they contain. Here, $A_1$ has many entities, $4/10 = 40\%$ of them, but these entities don’t possess very much of the $z$-trait. So, they will drag the average down much more than, say, $A_2$, which has a modest amount of $z$-trait and only comprises $1/10 = 10\%$ of the total population. $A_3$ will surely raise the average, but not by too much, since it only contains 2 entities, 20% of the population. $A_4$, on the other hand, contains a sizeable 30% of the total population, but the average amount of $z$-trait in $A_4$ seems neither large nor small. Hence, adding $A_4$ to the calculation is unlikely to significantly raise or lower the running average. Letting $\bar{A_j}$ denote the average level of $z$-trait in subpopulation $j \in \{1, 2, 3, 4\}$, the partition theorem tells us that the average (i.e. the expected value) of the whole population will be:

$$\frac{4}{10} \bar{A_1} + \frac{1}{10} \bar{A_2} + \frac{2}{10} \bar{A_3} + \frac{3}{10} \bar{A_4}.$$ 

The reasoning applied here is exactly the reasoning that underlies

---

A Appendix

Figure A.1: The Partition Theorem Illustrated by a Truly Great Artist

See any probability textbook for an explanation. A particular statement, couched in terms of expected values, rather than probabilities, can be found in Grimmett and Walsh (2014, 34).
the partition theorem, and if you understood this reasoning, then you are (at least) very close to understanding why $\bar{z}_0 = \frac{1}{n} \sum_{i=1}^{n} \omega_i z_i'$. 

**Deriving $\Delta \bar{z}$**

Again, what we’re after is a simple expression that describes the change in the average level of the $z$-trait. We will denote this quantity $\Delta \bar{z}$. Now, to begin this derivation, we simply note the fact that $\Delta \bar{z}$ must be equal to the difference between the average in the new $P$-population, and the old $P$-population.

\[
\Delta \bar{z} = \bar{z}_0 - \bar{z} = \frac{1}{n} \sum_{i=1}^{n} \omega_i z_i' - \frac{1}{n} \sum_{i=1}^{n} z_i
\]

Recall, \( z_i' = \Delta z_i + z_i \)

Hence,

\[
\Delta \bar{z} = \frac{1}{n} \sum_{i=1}^{n} \omega_i (\Delta z_i + z_i) - \frac{1}{n} \sum_{i=1}^{n} z_i
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \omega_i z_i - \frac{1}{n} \sum_{i=1}^{n} z_i + \frac{1}{n} \sum_{i=1}^{n} \omega_i \Delta z_i
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \omega_i z_i - \mathbb{E}[\omega] \frac{1}{n} \sum_{i=1}^{n} z_i + \frac{1}{n} \sum_{i=1}^{n} \omega_i \Delta z_i
\]

\[
= \mathbb{E}[\omega z] - \mathbb{E}[\omega] \mathbb{E}[z] + \mathbb{E}[\omega \Delta z]
\]

Therefore,

\[
\Delta \bar{z} = \mathbb{Cov}(\omega, z) + \mathbb{E}[\omega \Delta z].
\]

(Price Eq.)

For some reason, presumably mathematical tractability, biologists typically prefer to multiply Price Eq. through by $\bar{w}$, yielding an equivalent, but more familiar, statement of the Price equation:

\[
\bar{w} \Delta \bar{z} = \mathbb{Cov}(w, z) + \mathbb{E}[w \Delta z]
\]

(A.2)

Notice that the only real change is that $\omega$ has been replaced with $w$. This is just because $\omega_i = \bar{w}_i/\bar{w}$, and we multiplied through by $\bar{w}$. The mathematical details here are both simple and unenlightening, so I will spare the reader. To obtain A.2, simply multiply equation (*) by $\bar{w}$ and simplify.\(^1\)

---

\(^1\) To see a derivation of (A.2) instead of (Price Eq.), see Okasha (2006, ch. 1).
A.2 Deriving the Multilevel Price Equation

The key idea of the multilevel version of the Price equation is to partition selective pressures into two categories: one between “collectives” and one between “particles,” i.e. the elements which make up the collectives. Building from equation A.2, there are two ways of arriving at the multilevel form of the Price equation. The top-down approach relies on the neat recursive trick of inserting the Price equation into itself, but indexing the inserted version to a lower level of selection. The bottom-up approach also relies on a mathematical trick, that of decomposing covariance into a within-partition term and a between-partition term.

Top-Down

Recall equation A.2:

\[ \bar{\Delta} z = \text{Cov}(w, z) + \mathbb{E}[w \Delta z]. \]

The covariance and expectation terms of this equation are taken over entities \( i \in P \). In this top-down version of the derivation, we will assume that each entity \( i \) is itself a collective. That is, each \( i \in P \) is itself made up of adaptive particles. We might think of each \( i \) as a group made up of individual organisms, or as an organism made up of genes which are sometimes capable of manipulating the meiotic process so as to increase their own spread, often at the expense of the progeny-organisms (meiotic drive).

To formalize this new situation, we will need some additional notation. Let \( j \in i \) be particles in group \( i \). For ease of mathematics, and without loss of generality, we assume that all groups \( i \) are of the same size. In this context, we will slightly abuse notation to draw an important distinction: \( \text{Cov}_i(w, z) \) will represent the covariance between group-fitness (i.e. the average number of offspring produced by the particles in \( i \)) and group-trait level (i.e. the average level of \( z \) possessed by the particles in \( i \)). \( \text{Cov}_i(w, z) \), on the other hand, will represent the covariance between particle-fitness (i.e. the number of offspring produced by each \( j \in i \)) and group-trait level (i.e. the level of \( z \) possessed by the \( j \in i \) within some group \( i \)). Similarly, \( \mathbb{E}_i[w \Delta z] \) will represent the expectation of the product of group-fitness (i.e. the average number of offspring produced by the particles in \( i \) multiplied by the change in group-trait level (i.e. the change in average level of \( z \) possessed by the particles in \( i \) between generations)). \( \mathbb{E}_i[w \Delta z] \), on the other hand, will represent the expectation of particle-fitness (i.e. the number of offspring produced by each \( j \in i \)) multiplied by the change in the individuals’ levels of \( z \)-trait (i.e. the level of \( z \) possessed by the \( j \in i \)). More briefly, when the expectation is indexed by \( i \), the expectation sums across the groups \( i \) that partition \( P \). When the expectation is indexed by \( j \), the expectation sums across particles \( j \) that comprise some specified group \( i \). And similarly for covariance.

We will derive the following:

**Theorem A.2.1** In the multilevel context, equation A.2 is equivalent to the
following:
\[ \bar{\omega} \Delta \bar{z} = \text{Cov}_t(w, z) + E_i \left[ \text{Cov}_j(w, z) + E_j [w \Delta z] \right] \tag{A.3} \]

Proof. Starting with equation A.2 and inserting our new notation . . .
\[ \bar{\omega} \Delta \bar{z} = \text{Cov}(w, z) + E [w \Delta z] \]
\[ = \text{Cov}_t(w, z) + E_j [w \Delta z] \]

That is, we apply the Price equation at the level of collectives, made up of lower level particles that are not yet represented in the equation. We are moving in a "top-down" direction. Let \( \text{I}_i \) be the number of particles in each group and \( n \) be the number of groups. To represent the lower-level particles explicitly, we note that . . .
\[ E_8 [w \Delta z] = E_8 \times E_9 \left[ \bar{\omega}_j \Delta \bar{z}_j \right] \]
\[ = \frac{1}{n} \sum_{j=1}^{n} \left( \frac{1}{m} \sum_{j=1}^{m} \bar{\omega}_j \frac{1}{m} \sum_{j=1}^{m} \Delta z_j \right) \]
\[ = \frac{1}{n} \sum_{j=1}^{n} (\bar{\omega}_j / \Delta z_j) \]
\[ = E_9 \left[ \bar{\omega}_j \Delta \bar{z}_j \right] \]

Here, we note that the expression within the expectation identical to the left-hand side of equation A.2, except that it is indexed to the lower level of particles \( j \), rather than the level of collectives \( i \). Accordingly, we can pull the clever move I alluded to above, of recursively inserting the Price equation into itself:
\[ E_i [w \Delta z] = E_i \left[ \bar{\omega}_j \Delta \bar{z}_j \right] \]
\[ = E_i \left[ \text{Cov}_j(w, z) + E_j [w \Delta z] \right] \]

We now simply insert (j-level) into the expression (i-level) to yield our result:
\[ \bar{\omega} \Delta \bar{z} = \text{Cov}_t(w, z) + E_i \left[ \text{Cov}_j(w, z) + E_j [w \Delta z] \right] \tag{A.3} \]

Before re-deriving (A.3) from the bottom up, consider its interpretation. This equation partitions the selective forces into two levels: the i-level of collectives, represented by \( \text{Cov}_t(w, z) \) and the j-level of particles, represented by \( E_i \left[ \text{Cov}_j(w, z) + E_j [w \Delta z] \right] \). What would happen if we eliminated evolutionary pressures at the particle level by assuming that all particles “breed true” (\( z' = z \) and hence \( E_j [w \Delta z] = 0 \)) and that they all have the same fitness (so that \( z_j \) is not correlated with \( w_j \))? We are left only with \( \text{Cov}_t(w, z) \), which measures how collective-level fitness (the average fitness of a collective’s particles) corresponds to collective-level \( z \)-trait (the average \( z \)-trait level of a collective’s particles). In other words, we are left with selection at the level of collectives. Alternatively, suppose that there is no collective-level selection. That is, a higher average level of the \( z \)-trait within a group does not correspond to greater proliferation of its members relative to other groups. Perhaps all collectives have the
same fitness, or, for whatever other reason, $w_i$ is not correlated with $z_i$. Then, of course, $\text{Cov}(w_i, z_i) = 0$ and we are left only with the term $E_i [\text{Cov}_i (w, z) + E_i [w \Delta z]]$, which takes the average across groups of the evolutionary outcomes that occur within groups at the particle level. In other words, the evolutionary process is entirely determined by selection and transmission bias at the particle-level.

One more point, which will serve as a transition to bottom-up thinking, bears mentioning. The term $E_i [\text{Cov}_i (w, z) + E_i [w \Delta z]]$ has been described in two distinct ways: (1) as selection and transmission bias when seen from the particle-level, and (2) as the transmission bias when seen from the collective level. The Price equation thus reveals an interesting fact about multilevel selection: transmission bias at level $i$ is an evolutionary process unto itself at level $(i-1)$. To take a concrete example, if we observe the preferential transmission of a particular allele in a human population, this can be understood either as a sort of transmission bias at the phenotypic level ($z' > z_i$ and so $E [w_i \Delta z_i] > 0$), or we can consider it to be a process of meiotic drive, an evolutionary process unto itself, taking place at the level of competing genes. As we will see shortly, starting at either of these levels, we can build upwards to consider how lower-level evolutionary processes determine evolution at higher levels. Evolving systems, at least in the Price framework, can be seen as a vast series of nested and inter-determining levels of selection.

**Bottom-Up**

Just as the top-down derivation employed a clever recursive trick, so this bottom-up derivation utilizes another tool in the arsenal of mathematical analysis: decomposing covariance into a within- and a between-group component. To demonstrate this decomposition I rely on graphics and intuition. A fully rigorous mathematical approach can be found in (Wade 1985, 62-3).

Consider the covariance term we wish to decompose: $\text{Cov}(w, z)$. Our decomposition involves separating an observed dispersion into two components: one that is “explainable” by the group to which the particles belong and another explainable by the dispersion within the group. Suppose we have a total population $P$ consisting of 9 particles, all of whom have varying levels of $z$-trait. There are 9 total particles, $j \in \{1, \ldots, 9\}$, but considered as members of a group, each particle will be indexed with an $i \in \{1, 2, 3\}$. We thus have three group, $k \in \{A, B, C\}$. 

The sum is often multiplied by \( \frac{1}{n} \) to give us an average, denoted \( s^2 \), rather than a total that strictly increases as we add more particles. When used for estimation purposes, as in many statistics textbooks, the sum is divided by \( n - 1 \) rather than \( n \) for somewhat esoteric reasons involving the desire for an unbiased estimator.

Each circle marker in this graph represents a particle \( j \) in the population \( P \). In this figure, the fact that they are grouped into three distinct collectives is irrelevant. We are considering the total dispersion of the particles, measured by the sum of their distance from the population mean, represented by the horizontal green line. The center particle, directly above ‘B’, shows how we measure this quantity: for each particle, we draw a line like the one connecting the center particle to the mean. Then we sum up this different across all particles. Now, this would be a poor measure of dispersion, since extremely above-average and extremely below average particles would cancel out, making the dispersion seem small when it is really large. For that reason, these difference are typically squared to yield the formula for mean-squared distance:

\[
MSD = \sum_{j=1}^{n}(z_j - \bar{z})^2,
\]

where \( z_j \) is as above, \( \bar{z} \) is the population average z-level (here, 4.66).

Our formula will be rather different, however, since we are not interested in variance, but in covariance. So, we must construct a second graph, similar to the first, except on the y-axis we have fitness \( w \), instead of \( z \). Then, instead of MSD, we will calculate:

\[
\text{Cov}(w, z) = \frac{1}{n} \sum_{j=1}^{n}(w_j - \bar{w})(z_j - \bar{z}).
\]

The second graph corresponds to the first term in the covariance expression, \( (w_j - \bar{w})(z_j - \bar{z}) \), which measures our fitness dispersion:
Again, we calculate the dispersion by subtracting each $j$'s value from the population mean. Our expression for covariance now conveys useful information: if $w_j$ tends to move in the same direction as $z_j$, then $\text{Cov}(w, z)$ will be fairly large. If, on the other hand, they tend to move in opposite directions, then $\text{Cov}(w, z)$ will be negative. If they exhibit little to no correspondence, then $\text{Cov}(w, z)$ will be close to 0.

Notice, further, that we have represented $z$ as a fairly altruistic trait: within groups, those with higher $z$-level have fitness that is lower than the group average. Consider for example, $j = 2$ with a high $z_2 = 8$, but a fitness level $w_2$ that is lower than either of its two collective members in A. Nevertheless, group A as a whole does quite well. Even the highly self-sacrificial particle $j = 2$ outperforms all members of the selfish group C.

Now, what the decomposition technique shows is that we can break up these dispersion-measures into within- and between-group components. This is easily visualized. First consider between-collective dispersion:
Now the centered vertical line measures the distance between the B-collective mean and the population average, representing the center particle indirectly, only through its influence on the collective mean (the thick blue bar). In practice, we draw a similar graph for fitness \( w \), calculating the distance between each collective’s mean and the population mean. Then we calculate the between-group covariance of \( z \)-level and \( w \), denoting this value \( \text{Cov}_1(w, z) \). We also run through a similar process to calculate within-group covariance between \( z \)-level and \( w \).

Here, we add up the differences between particles’ \( z \)-levels and the collective average, rather than the population average. Hence, we draw a line from the center particle to 5.33, its within-collective average, rather than to 4.66, the total population average. Again, we do something similar to find the dispersion of within-group fitness \( w \). Now, clearly, differences within the group account for much of the total population dispersion. But, just as clearly, they do not account for all of it, because the within-collective averages are (almost by definition) closer to their within-group particles than they are to the whole population-wide gamut of particles. The rest of the population dispersion is captured by the dispersion of collective means from the population mean, visualized above in figure A.4. The claim of the decomposition technique is that the total population dispersion can be captured by summing the average within-collective covariance and the between-collective covariance.

With this intuition, let’s briefly formalize the decomposition claim. If we let \( i \) index particles within collectives \( k \), then \( w_{ik} \) is the \( i \)th particle in the \( k \)th group, with \( i \in \{1, 2, 3\} \) and \( k \in \{A, B, C\} \). For example, \( w_{2B} \) denotes the fitness of the center particle, i.e. the second particle in the B group. Similarly, \( z_{ik} \) is the \( z \)-level of the \( i \)th particle in the \( k \)th group. If we write simply \( w_k \), then we are denoting the within-collective mean of \( k \). As above, \( \text{Cov}_1(w, z) \) denotes the covariance between the within-collective average fitness and the within collective average \( z \)-level, expanded this is: \( \text{Cov}_1(w, z) = \frac{1}{3} \sum_{k \in A} (w_k - \bar{w})(z_k - \bar{z}) \). On the other hand, we will use \( \text{Cov}_1 \) to denote a covariance within a group \( k \), so that \( \text{Cov}_1(w, z) = 4 \sum_{i=1}^{3} (w_{ik} - \bar{w}_k)(z_{ik} - \bar{z}_k) \). With this notation, we can
formally spell out the intuitive claim that total population dispersion is decomposable into between- and within-group components:

\[ \text{Cov}(w, z) = \text{Cov}_k(w, z) + \mathbb{E}_k[\text{Cov}_i(w, z)] \]  

(A.4)

The population-covariance between fitness \( w \) and \( z \)-trait is equal to the sum of (i) the covariance between collective fitness \( w_k \) and collective \( z \)-trait (i.e. the averages within the collective) and (ii) the mean of the within-collective covariances between fitness \( w \) and \( z \)-trait.

Armed with equation A.4, the bottom-up derivation of equation A.3 is quite simple.

**Theorem A.2.2** In the multilevel context, equation A.2 is equivalent to the following:

\[ \bar{\omega} \Delta \bar{z} = \text{Cov}_k(w, z) + \mathbb{E}_k[\text{Cov}_i(w, z) + \mathbb{E}_i[w \Delta z]] \]

**Proof.** Recall equation A.2:

\[ \bar{\omega} \Delta \bar{z} = \text{Cov}(w, z) + \mathbb{E}[w \Delta z] \]

\[ = \text{Cov}_k(w, z) + \mathbb{E}_k[\text{Cov}_i(w, z)] + \mathbb{E}[w \Delta z] \]

An easily derivable fact about expected value is that, for any partition of the population \( P \), the expectation of the total population is equal to the expected value of the sum of the expected values of the partitions. Hence,

\[ \mathbb{E}[w \Delta z] = \mathbb{E}_k[\mathbb{E}_i[w \Delta z]] = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i=1}^{n_k} (w_{ik} \Delta z_{ik}) \]

where \( K \) is the total number of collectives (elements in the partition), \( n_k \) is the total number of particles in collective \( K \), \( w_{ik} \) is the fitness level of the \( i \)th particle in the \( k \)th collective, and \( z_{ik} \) is the \( z \)-level of the \( i \)th particle in the \( k \)th collective.

Inserting this equation into the above expression...

\[ \bar{\omega} \Delta \bar{z} = \text{Cov}_k(w, z) + \mathbb{E}_k[\text{Cov}_i(w, z)] + \mathbb{E}_k[\mathbb{E}_i[w \Delta z]] \]

\[ = \text{Cov}_k(w, z) + \mathbb{E}_k[\text{Cov}_i(w, z) + \mathbb{E}_i[w \Delta z]] \]

In contrast to the top-down derivation, this one began with particles, not considering whether these particles are themselves collectives comprised of lower level particles. They may be. We could then apply the same top-down derivation to these particles, identifying them as intermediary collectives, made up of particles but also constituting particles for higher-level collectives. Evolutionary systems are thus seen to be nested processes, teeming with activity at each level, an activity that often expresses itself through emergent coherence and unity at higher levels.
A.3 Henrich’s Defense of Cultural Group Selection

Recall the basic form of the Price Equation:

\[ \Delta \bar{z} = \text{Cov}(w_i, z_i) + \text{E}[w_i \Delta z] \]  \hfill (A.5)

Now, let us consider fitness of individual \( i \), \( w_i \), as depending partly on the trait \( z_i \) and partly on the average level of the altruistic trait within the group, denoted \( z_j \). Given that it is an altruistic trait, fitness will be undermined by the possession of this trait, but enhanced by others possessing it. And since \( i \)’s possession of the trait will be correlated with the group’s possession of this trait, we need a way to separate out these two effects. A simple regression equation can do this for us. Let \( \beta_{w_i,z_i} \) be the partial regression coefficient for \( w_i \) on \( x_i \). That is, \( \beta_{w_i,z_i} \) represents the strength of the effect that changing \( z_i \) has on \( w_i \) holding \( z_j \) constant. Similarly, \( \beta_{w_i,x_i} \) will represent the partial regression coefficient for \( w_i \) on \( x_i \). That is, \( \beta_{w_i,x_i} \) represents the strength of the effect that changing \( z_j \) has on \( w_i \) holding \( z_i \) constant. As is standard, we will also include the \( y \)-intercept term \( \beta_0 \) and an error term \( \epsilon \):

\[ w_i = \beta_0 + \beta_{w_i,z_i} x_i + \beta_{w_i,x_i} z_j + \epsilon \]  \hfill (A.6)

Now, take the expression in A.6 and substitute it into expression A.5. Let \( \Delta z = 0 \), since we are not interested in drift, but selection. So, setting \( \text{E}[w_i \Delta z] = 0 \) and substituting...

\[ \Delta \bar{z} = \text{Cov}(\beta_0 + \beta_{w_i,z_i} z_i + \beta_{w_i,x_i} z_j + \epsilon, z_i) \]  \hfill (A.7)

**Relevant Covariance Rules**

\[
\begin{align*}
\text{Cov}(X + Y, Z) &= \text{Cov}(X, Z) + \text{Cov}(Y, Z) \\
\text{Cov}(aX, Z) &= a\text{Cov}(X, Z) \\
\text{Cov}(X, X) &= V(X) \\
\text{Cov}(X, Y) &= \beta_{XY} V(X)
\end{align*}
\]

Applying covariance rules to equation (A.7), we get:

\[ \Delta \bar{z} = \beta_{w_i,z_i} V(z_i) + \beta_{w_i,z_j} \text{Cov}(z_j, z_i) \]

Because \( \text{Cov}(z_j, z_i) = \text{Cov}(z_i, z_j) = \beta_{z_j,z_i} V(z_i) \)...

\[ \Delta \bar{z} = \beta_{w_i,z_i} V(z_i) + \beta_{w_i,z_j} \beta_{z_j,z_i} V(z_i) \]

\[ = (\beta_{w_i,z_j} + \beta_{z_j,z_i} \beta_{w_i,z_i}) V(z_i) \]  \hfill (□)

To understand the meaning of this last expression, we can borrow from the analysis of frank1998foundations to reconstruct the complicated coefficient term in (□). Let the regression coefficient \( \beta_{w,z} \) denote exactly what we want to know: the total, non-decomposed, statistical (i.e.
not necessarily causal) effect of a trait $z_i$ on the organism’s fitness, $w_i$. Frank points out that this regression coefficient contains two separable components. One is the direct, causal effect of $z_i$ on $w_i$, ignoring any indirect effects that result from the group. We have been representing this component with the familiar partial regression coefficient $\beta_{w_i|z_i}$. The second component consists of (i) the statistical effect of trait $z_i$ on the group-level of trait $z$, denoted $z_j$, multiplied by (ii) the causal effect of the group level $z_j$ on an organism’s fitness $w_i$. We denote (i) with the regression coefficient $\beta_{w_i|z_j}$ and (ii) with the partial regression coefficient $\beta_{w_i|z_i; z_j}$. This can be visualized by the following diagram, adapted from (Frank 1998, 52):

![Diagram](image)

**Figure A.6:** Relating partial regression coefficients and fitness

As this diagram indicates...

$$\beta_{w_i|z_i} = \beta_{w_i|z_i} + \beta_{z_i|z_i} \beta_{w_i|z_i; z_i} \quad (\square)$$

Comparing $(\square)$ and $(\Box)$, we see that the right hand side of $(\Box)$ is equal to the coefficient in front of $V(z_i)$ in $(\square)$. This is fantastic, because we now have some intuitive sense of what the the expression $(\Box)$ actually signifies. The adaptiveness, or tendency to increase, of a trait $z$ is regulated partly by the effect it has on the organism’s fitness, but also partly by the product of (i) how likely an organism with that trait is to be surrounded by other organisms with that trait and (ii) the impact on the organism of being in a group with the trait $z$. It’s nice to understand what the mathematics means, but we should also ask what we can do with it.

Henrich is able to get some serious mileage out of the expression $(\Box)$. Since $V(z_i) \geq 0$ as a mathematical fact, this implies that a trait will be selected for, i.e. $\Delta \bar{w} \Delta \bar{z} \geq 0$, only if we have:

$$\beta_{w_i|z_i; z_j} + \beta_{z_i|z_i} \beta_{w_i|z_i; z_i} > 0 \quad (***)$$

If we assume we are talking about an altruistic trait, then we know the following facts:

$$\beta_{w_i|z_i; z_j} < 0$$
$$\beta_{w_i|z_i; z_i} > 0$$

* Again, remember that, although we are treating $z$ as a binary trait for ease of expression, the same reasoning applies to a continuous trait.
Given these two facts, our prediction of whether an altruistic trait will spread depends upon a crucial structural feature of the population. In order to ensure that condition (***) is satisfied, we would like $\beta_{x,z}$ to be large and positive. In other words, going back to the intuitive meaning of the expression, an altruistic trait is more likely to be selected when altruists are capable of bunching together. This point is of fundamental importance for understanding multilevel selection and the evolution of altruism, so it bears repeating: in order for an altruistic trait to evolve (through selection), altruists must have some mechanism(s) for excluding egoists from their network of interaction or, equivalently, of converting egoists within their network into altruists.
References


Campbell, Donald T. "Factors relevant to the validity of experiments in social settings." Psychological bulletin 54, no. 4 (1957): 297.


Green, S. 2005. “Can Biological Complexity Be Reverse Engineered?” In Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 53: 73-83


Ostrom, E. and H., Nagendra. 2006. Insights on linking forests, trees, and people from the air, on the ground, and in the laboratory. Proceedings of the national Academy of sciences, 103(51): 19224–19231.


