Developments in Complex Systems Science with Applications to Political Systems and Pandemic Response

by

Alexander F. Siegenfeld
S.B., Massachusetts Institute of Technology

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 2022

Author .....
Department of Physics
April 22, 2022

Certified by .....
Yaneer Bar-Yam
Professor and President, New England Complex Systems Institute
Thesis Supervisor

Certified by .....
Deb Roy
Professor of Media Arts and Sciences
Thesis Supervisor

Certified by .....
Mehran Kardar
Francis Friedman Professor of Physics
Thesis Supervisor

Accepted by .....
Deepto Charkrabarty
Associate Department Head

Developments in Complex Systems Science with Applications to Political Systems and Pandemic Response

by

Alexander F. Siegenfeld

Submitted to the Department of Physics
on April 22, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The standard assumptions that underlie most conceptual and quantitative frameworks do not hold for many complex physical, biological, and social systems. Complex systems science clarifies when and why such assumptions fail and provides alternative frameworks for understanding the properties of complex systems. We review some of the basic principles of complex systems science and provide a mathematical formalism for complexity profiles. We also illustrate general modeling principles using examples from pandemic response, including an illustration of how pandemics can be stably eliminated with a combination of social distancing measures and travel restrictions and how bad science led to bad policy regarding the use of face masks. Applications to democratic elections are also described. We define the concepts of negative representation and electoral instability, demonstrating that United States' presidential elections underwent a transition from a stable to an unstable regime in the 1970s and have since become increasingly unstable. We also consider the implications of geographic political polarization on multi-scale electoral systems.

Thesis Supervisor: Yaneer Bar-Yam
Title: Professor and President, New England Complex Systems Institute

Thesis Supervisor: Deb Roy
Title: Professor of Media Arts and Sciences

Thesis Supervisor: Mehran Kardar
Title: Francis Friedman Professor of Physics

Acknowledgments

I am very fortunate to have had many mentors and collaborators throughout my PhD. I am indebted to my advisors Prof. Yaneer Bar-Yam, Prof. Deb Roy, and Prof. Mehran Kardar for their invaluable mentorship and for everything I have learned from them. I am also grateful to my thesis committee members Prof. Edmund Bertschinger and Prof. Jeff Gore for their very helpful feedback and guidance over the past few years. I would also like to thank my collaborators William Brannon, Pratyush Kollepara, Blake Elias, Sihao Huang, Robin Na, Alina Harbuzova, Krit Boonsiriseth, Robi Bhattacharjee, Alex Zhu, Prof. Nassim Nicholas Taleb, and Prof. Andrew Gelman.

This work could not have occurred without financial support: I gratefully acknowledge the National Science Foundation Graduate Research Fellowship Program (under Grant No. 1122374), the Hertz Foundation, the Long-Term Future Fund, the MIT Department of Physics, the New England Complex Systems Institute, and the MIT Center for Constructive Communication. The financial and community support of the Hertz Foundation in particular was instrumental in my decision to choose this unorthodox research direction.

I would like to thank the research teams and staff at the New England Complex Systems Institute and the MIT Center for Constructive Communication, as well as the MIT Department of Physics staff, for all of their support.

Finally, I am beyond grateful to all of my family, friends, teachers, and mentors, who have shaped and supported me throughout my life. I am especially indebted to my siblings, Allison Siegenfeld and Mikoto Araki-Siegenfeld, for their unconditional love and support and, of course, to my parents, Barbara and Michael Siegenfeld, and stepmother, Kasumi Araki-Siegenfeld, for the values, education, love, and joy with which they raised me.

Contents

Chapter 1

Introduction

How can we scientifically approach the study of complex systems—physical, biological, and social? Empirical studies, while useful, are by themselves insufficient, since all experiments require a theoretical framework in which they can be interpreted (whether that framework be qualitative or quantitative, implicit or explicit, wholly prior to or arising in part from the data, etc.) [205]. While many such frameworks exist for understanding particular components or aspects of systems, the standard assumptions that underlie most quantitative studies often do not hold for systems as a whole, resulting in a mischaracterization of the causes and consequences of large-scale behavior.[1]

We first provide an introduction to complex systems science, demonstrating a few of its applications and its capacity to help us make more effective decisions in the complex systems of our world. Mathematical formalisms for the concepts described in section 1.1 are then given in chapter 2. Principles for how to model real-world systems are illustrated using case studies in modeling pandemics in chapter 3, with a specific multi-scale model for pandemics given in chapter 4. Chapter 5 describes how bad science led to bad policy regarding the use of masks against COVID-19. Chapters 6 and 7 describe applications to political systems.

[1]Most of this chapter is a slightly modified version of the preprint `https://arxiv.org/abs/1912.05088`, published as: Alexander F. Siegenfeld and Yaneer Bar-Yam. An introduction to complex systems science and its applications. *Complexity* 6105872 (2020).

Chapter 8 concludes with an outlook on the field of complex systems science and future research directions.

This chapter focuses on the general properties of complex systems. Section 1.1 introduces key concepts, including complexity profiles, the tradeoff between efficiency and adaptability, and the necessity of matching the complexity of systems to that of their environments. Section 1.2 considers the analysis of complex systems, attending to the oft-neglected question of when standard assumptions do and—more importantly—do not apply. Section 1.3 discusses principles for effectively intervening in complex systems given that their full descriptions are often beyond the limits of human comprehension. Section 1.4 provides further reading. Section 1.5 summarizes this chapter.

1.1 Basic Principles of Complex Systems Science

1.1.1 Why complex systems science?

Complex systems science considers systems with many components. These systems could be physical, biological, or social. Given this diversity of systems, it may seem strange to study them all under one framework. But while most scientific disciplines tend to focus on the components themselves, complex systems science focuses on how the components within a system are related to one another [40]. For instance, while most academic disciplines would group the systems in fig. 1-1 by column, complex systems science groups them by row.

Systems may differ from each other not because of differences in their parts but because of differences in how these parts depend on and affect one another. For example, steam and ice are composed of identical water molecules but, due to differences in the interactions between the molecules, have very different properties. Conversely, all gasses share many behaviors in common despite differences in their constituent molecules. The same holds for solids and liquids. The behaviors that distinguish solids from liquids from gasses are examples of *emergence*: they cannot be determined from a system's parts individually. Fluid turbulence, as one might observe in a flowing river, is an example of how the relationships between parts

can give rise to emergent large-scale behaviors and patterns that are self-organized, meaning that they arise not from some external or centralized control but rather autonomously from the interactions between the system components [16, 194, 127, 106, 160]. Other examples of self-organized behaviors include the spontaneous formation of conversation groups at a party, the allocation of goods in a decentralized economy, the evolution of ecosystems, and the flocking of birds. Such large-scale behaviors and patterns cannot be determined by examining each system part in isolation. By instead considering general properties of systems as wholes, complex systems science provides an interdisciplinary scientific framework that allows for the discovery of new ideas, applications, and connections.

A full description of all the small-scale details of even relatively simple systems is impossible; therefore sound analyses must describe only those properties of systems that do not depend on all these details. That such properties exist is due to *universality*, a phenomenon that will be discussed in section 1.2. Statistical physics provides an underlying insight that allows for the discovery of such properties: namely, that while attempting to characterize the behavior of a particular state of a system (e.g. a gas) may be entirely intractable, characterizing the set of all possible states of the system may not only be tractable but may also provide us with a model of the relevant information (e.g. the pressure, temperature, density, compressibility, etc.). In other words, taking a step back and considering the *space of possible behaviors* provides a powerful analytical lens that can be applied not only to physical systems but also to biological and social ones.

1.1.2 What is complexity?

We define the complexity of a behavior as equal to the length of its description. The length of a description of a particular system's behavior depends on the number of possible behaviors that system could exhibit [101]. For example, a light bulb that has two possible states—either on or off—can be described by a single bit: 0 or 1. Two bits can describe four different behaviors (00, 01, 10, or 11), three bits can describe eight behaviors, and so on. Mathematically, we can write $C = \log_2 N$, where C is the complexity of a system and N is

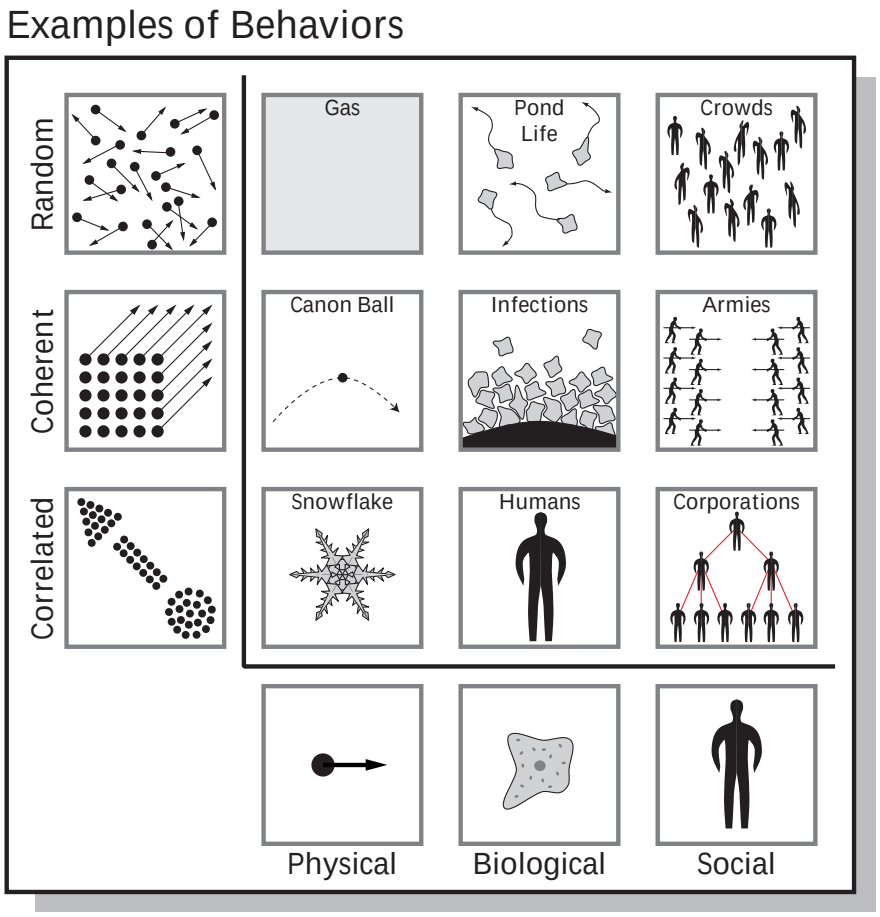


Figure 1-1: From ref. [41]. Each column contains three examples of systems consisting of the same components (from left to right: molecules, cells, people) but with different relations between them. Each row contains systems representing a certain kind of relationship between components. For random systems, the behavior of each component is independent from the behavior of all other components. For coherent systems, all components exhibit the same behavior; for example, the behavior (location, orientation, and velocity) of one part of the cannonball completely determines the behavior of the other parts. Correlated systems lie between these two extremes, such that the behaviors of the system's components do depend on one another, but not so strongly that every component acts in the same way; for example, the shape of one part of a snowflake is correlated with but does not completely determine the shape of the other parts. (Implicit in these descriptions is the necessity of specifying the set of behaviors under consideration, as discussed in section 1.1.2.)

its number of possible behaviors,[2] but for our purposes here, it is sufficient to state that the greater the number of possible behaviors, the greater the complexity.

It is important to note that one must carefully define the space of possible behaviors. For instance, if we are interested in a light bulb already in a socket, the light bulb has two possible behaviors, as above, but if we are instead interested in the complexity of building a light bulb, the space of possible behaviors might include all of the ways in which its parts could be arranged. As another example, consider programming a computer to correctly answer a multiple-choice question with four choices. At first glance, this task is very simple: since there are four possible behaviors, only two bits are required. Nonetheless, we have the sense that programming a computer to score perfectly on a multiple-choice test would be quite difficult. This apparent paradox is resolved, however, when we recognize that such a task is difficult only because we do not *a priori* know what questions will be on the test, and thus the true task is to be able to correctly answer *any* multiple-choice question. This task is quite complex, given the large number of possible ways the program could respond to a string of arbitrary multiple-choice questions.

1.1.3 Complexity and scale

Consider a human, and then consider a gas containing the very same molecules that are in the human but in no particular arrangement. Which system is more complex? The gas possesses a greater number of possible arrangements of the molecules (i.e. has more entropy, or disorder), and thus would take longer to describe at a microscopic level. However, when we think of a complex system, we think of the behaviors arising from the ordered arrangement of molecules in a human, not the behaviors arising from the maximally disordered arrangement of molecules in a gas. It therefore may be tempting to conclude that complex systems are those with reduced disorder. But the systems with the least disorder are those in which all

[2]Technically, $\log_2 N$ is actually an upper bound for the system's complexity since if some behaviors are more likely than others, the average length of the system's description can be reduced by using shorter descriptions for the more common behaviors and longer descriptions for the less common ones. (Lossless compression algorithms rely on this logic.)

components exhibit the same behavior (coherent systems in fig. 1-1), and such behavior is easy to describe and thus not intuitively complex.

To resolve this apparent paradox, we must consider that the length of a system's description depends on the level of detail used to describe it. Thus, complexity depends on scale. On a microscopic scale, it really is more difficult to describe the positions and velocities of all the molecules of the gas than it is to do the same for all the molecules of the human. But at the scale of human perception, the behaviors of a gas are determined by its temperature and pressure, while the behaviors of a human remain quite complex. Entropy corresponds to the amount of complexity at the smallest scale, but characterizing a system requires understanding its complexity across multiple scales. A system's *complexity profile* is a plot of the system's complexity as a function of scale [12]. In the examples below, scale will be taken to be length, but fundamentally, the scale of a behavior is equal to the number of coordinated components involved in the behavior, for which physical length is a proxy. A gas is very simple at the scale of human perception because at this scale, only behaviors involving trillions of molecules are relevant, and there are relatively few distinguishable behaviors of a gas involving so many molecules.

As shown in fig. 1-2, random, coherent, and correlated systems (see fig. 1-1) have qualitatively different complexity profiles. Random systems have the most complexity at the smallest scale (finest granularity/most detail), but the amount of complexity rapidly drops off as the scale is increased and the random behaviors of the individual components are averaged out. A coherent system has the same amount of complexity at small scales as it does at larger scales because describing the overall behavior of the system (e.g. the position and velocity of a cannonball) also describes the behavior of all the components (e.g. the positions and velocities of all the atoms). Note that complexity tends to increase (or remain the same) as the scale decreases, since looking at a system in more detail (while still including the whole system in the description) tends to yield more information. For a correlated system, various behaviors occur at various scales, and so the complexity gradually increases as one examines the system in greater and greater detail. For instance, from very far away a human, being

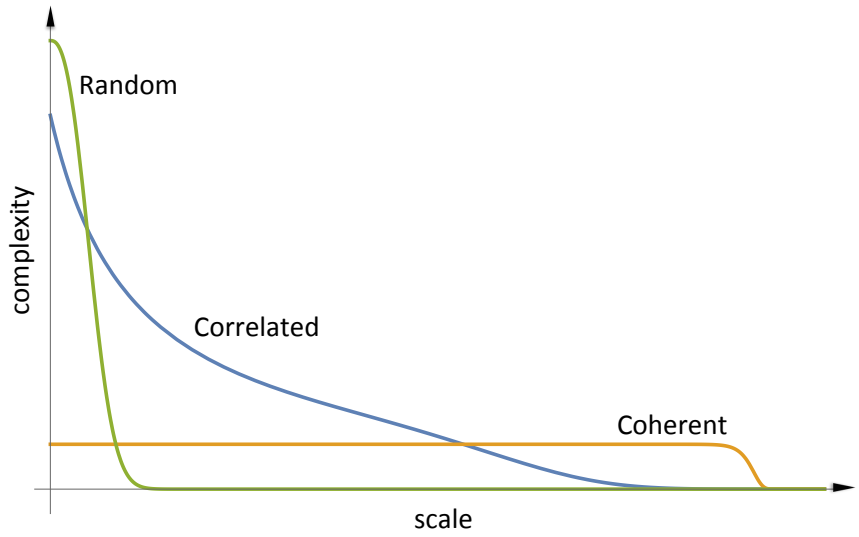


Figure 1-2: Representative complexity profiles for random, coherent, and correlated systems (see fig. 1-1). Any given system may have aspects of each at various scales.

barely visible, has very little complexity. As the level of detail is gradually increased, the description will first include the overall position and velocity of the human, and then the positions and velocities of each limb, followed by the movement of hands, fingers, facial expressions, as well as words that the human may be saying. Continuing to greater levels of detail, the organs and then tissues and patterns within the human brain become relevant, and eventually so do the individual cells. At scales smaller than that of a cell, complexity further increases as one sees organelles (cellular substructures), followed by large molecules such as proteins and DNA, and then eventually smaller molecules and individual atoms. At each level, the length of the description grows longer. This incredible multi-scale structure with gradually increasing complexity is a defining characteristic of complex systems.

1.1.4 Tradeoffs between complexity and scale

The intuition that complex systems require order is not unfounded: for there to be complexity at larger scales, there must be behaviors involving the coordination of many smaller-scale

components. This coordination suppresses complexity at smaller scales because the behaviors of the smaller-scale components are now limited by the interdependencies between them. The tension between small-scale and large-scale complexity can be made precise: given a fixed set of components with a fixed set of potential individual behaviors, the area under the complexity profile will be constant, regardless of the interdependencies (or lack thereof) between the components.[3] Thus, for any system, there is a fundamental tradeoff between the number of behaviors a system can have and the scale of those behaviors.

For instance, consider a factory consisting of many workers [41]. The output of the factory can be characterized using a complexity profile (fig. 1-3). The number of different types of goods that the factory can produce at a given scale is a proxy for the factory's complexity at that scale, with the number of copies of the same type of good that the factory can produce in a given amount of time being a proxy for scale. The fundamental tradeoff is evident in the fact that if the factory wants to be able to churn out many copies of a single type of good in a short amount of time, it will have to coordinate all of its workers (perhaps having them work on an assembly line), thereby reducing their individual freedom to make many different kinds of goods. The factory's production would then have low complexity but at a large scale (e.g. churning out many identical Model-T Fords—"Any customer can have a car painted any color that he wants so long as it is black"). On the other hand, if the factory's employees work independently, they will be able to create many different types of products, but none at scale. Of course, a factory may be able to increase both the complexity and scale of its production by adding new machinery or more workers; the precise tradeoff between complexity and scale applies only when considering a fixed set of components with a fixed set of individual behaviors.[4]

[3]Formally, the sum of a system's complexity at each scale (i.e. the area under its complexity profile) will equal the total complexity of its components, i.e. the sum of each individual component's complexity [12]. The complexity of an individual component is related to the number of distinct behaviors of that component, as described in section 1.1.2.

[4]A subtle point to be made here is that introducing interactions between two parts of a system may in some cases increase the set of relevant individual behaviors of each part, thereby increasing the total area under the complexity profile. For example, if two people enter into communication with each other, the communication itself (e.g. speech) may now be a relevant behavior of each individual person that was not there before.

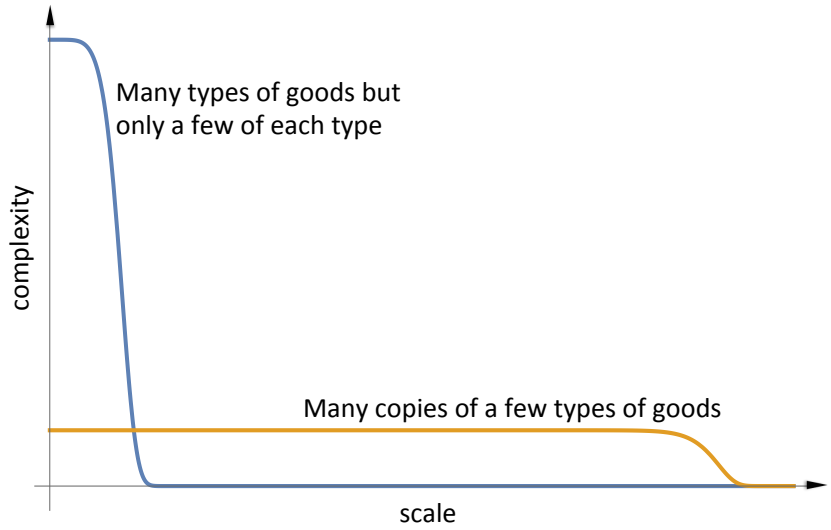


Figure 1-3: The complexity profile of a factory that can produce a large number of copies of a few types of goods, and the complexity profile of a factory that can produce many types of goods but not in large numbers. The number of copies of a good produced is a proxy for scale since, given a fixed technology, mass production requires larger-scale coordinated action in the factory (e.g. an assembly line), and the number of different types of goods that can be produced at a given scale is a proxy for the number of different possible behaviors of the factory—and thus its complexity—at that scale.

A corollary of the tradeoff between complexity and scale is the tradeoff between adaptability and efficiency [333, 202, 334, 347, 267, 260]. Adaptability arises when there are many possible actions happening in parallel that are mostly independent from one another, i.e. when the system has high complexity. Efficiency, on the other hand, arises when many parts of a system are all working in concert, so that the system can perform the task for which it was designed at the largest possible scale. Due to the tradeoff between complexity and scale, a system with more adaptability will have a complexity profile with greater complexity but predominantly at smaller scales, while a system with more efficiency will have a complexity profile with lower complexity but extending to larger scales. Thus, a very efficient system will, due to its necessarily lower complexity, not be as adaptable to unforeseen variations within itself or its environment, while a very adaptable system, designed to handle all sorts of shocks, will necessarily have to sacrifice some larger-scale behaviors. The Soviets thought they could have their cake and eat it, too: they originally believed that their economy would outperform capitalist ones because capitalist economies have so much waste related to multiple businesses competing to do the same thing [43, Chapter 16]. It would be far more efficient to coordinate all economic production. But in creating such large-scale economic structures, lower-scale complexity was sacrificed, resulting in a non-adaptive system. (Improperly regulated capitalist systems may also sacrifice redundancy and adaptability for efficiency, resulting in, for instance, excessive concentrations of market power, harmful feedback loops, and herd-like behaviors [116, 229, 217, 67, 164].)

Due to the tradeoff between complexity and scale, any mechanism that creates larger-scale complexity—whether market or government or otherwise—will necessarily reduce individual complexity. This is not to say that larger-scale complexity is always harmful; it is often worth trading some individual-level freedoms for larger-scale cooperation. When, then, is complexity at a particular scale desirable?

1.1.5 Why be complex?

A determination of when complexity is desirable is provided by the *Law of Requisite Variety* [22]: To be effective, a system must be at least as complex as the environmental behaviors to which it must differentially react. If a system must be able to provide a different response to each of 100 environmental possibilities and the system has only 10 possible actions, the system will not be effective. At the very least the system would need 100 possible actions, one for each scenario it could encounter. (The above condition is necessary but of course not sufficient; a system with sufficiently many actions may still not take the right actions in the right circumstances.) Note that the environment to which a system must react is itself also a system and will sometimes be referred to as such.

Since complexity is defined only with respect to a particular scale, we can refine the Law of Requisite Variety: To be effective, a system must match (or exceed) the complexity of the environmental behaviors to which it must differentially react at all scales for which these behaviors occur [12]. To illustrate this multi-scale version of the Law of Requisite Variety, we consider military conflict [42] (see fig. 1-4). Here, one military can be considered as the system, while the other military is part of the environment with which the system must interact. For two militaries of equal complexity, i.e. with the same number of behaviors, but with one military operating at a larger scale (e.g. two very tightly controlled armies, but with one army larger than the other), the larger-scale military will likely win. For two militaries of equal scale but unequal complexity (e.g. two equally sized and equally powered fleets, but with one being more maneuverable than the other), the higher-complexity military will likely win, since the high-complexity military has an action for every action of the lower-complexity military but not vice versa. When a military with high complexity at a smaller scale (e.g. a guerrilla force) conflicts with a military with larger-scale behavior but lower complexity (e.g. the U.S. army in Vietnam or the Soviet army in Afghanistan), the terrain, which constrains the scale of the conflict, plays an important role. In an open field, or in open waters, the military that has more complexity at the larger scales is favored, while in the jungle or in the mountains, higher complexity at smaller scales is favored.

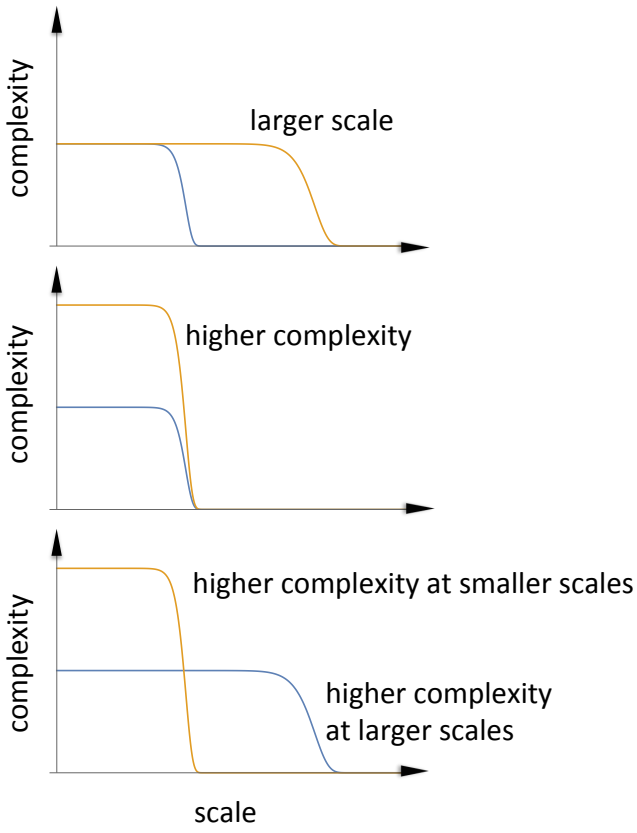


Figure 1-4: Schematic complexity profiles of militaries in conflict. Top: If two armies are operating with the same number of possible behaviors but at different scales, the larger-scale one is favored. Middle: If two armies are operating at the same scale but with different numbers of possible behaviors, the higher-complexity one is favored. Bottom: If two armies are operating at different scales and with different numbers of possible behaviors, which one is favored depends on the terrain (see text). Note that these profiles are simplified to highlight the key concepts; actual militaries operate at multiple scales. More generally, the top and middle graphs depict conflicts in which one army has at least as much complexity as the other at every scale.

As another example, healthcare involves both small-scale tasks with high overall complexity such as case management, as well as large-scale, lower complexity tasks, such as manufacturing and delivering vaccines [48]. (Delivering vaccines is lower complexity but higher scale because the same actions will be performed for nearly everyone.) Large-scale top-down organizations and initiatives are suited for large-scale, lower complexity tasks, but tasks like case management require health systems with a high degree of small-scale (i.e. local) complexity.

The eurozone provides a potential illustration of a multi-scale complexity mismatch. Fiscal policy is made predominantly at the scale of individual countries and thus has a higher complexity at the country scale but relatively little complexity at the scale of the entire eurozone, while monetary policy is made at the scale of the entire eurozone and thus has some complexity at the scale of the eurozone but lacks the ability to vary (i.e. lacks complexity) at the scale of individual countries. Many have argued that economic difficulties within the eurozone have arisen because this mismatch has precluded effective interactions between fiscal and monetary policy [82, 294, 9, 107, 125].

Problems arise not from too much or too little complexity (at any scale) per se but rather from mismatches between the complexities of a task to be performed and the complexities of the system performing that task.[5] Note that the system in one scenario may be the task/environment in another; for instance, the same complexity that helps a system interact with its environment may prevent its effective management by other systems. In none of the above examples have the complexity profiles been precisely calculated, nor have scales been precisely defined. Instead, proxies for scale are used and estimated comparisons of complexity made. Such an approach cannot yield precise results (indeed, no approach can, given the complexity a full description of such systems would require), but additional precision is not needed when even the approximate analysis reveals large mismatches in complexity.[6] (To

[5]Incidentally, human emotions appear to reflect this principle: we are bored when our environment is too simple and overwhelmed when it is too complex [108].

[6]Just as this analysis of the space of possible behaviors can be used even in the face of uncertainty regarding a system's precise mechanisms and outcomes, physicists can use the property of entropy (sometimes considering how quantities related to entropy change across scale) to classify phase transitions even when

remedy the diagnosed mismatches, more detailed analyses may be required.) While it may be tempting to attribute the problems arising from a complexity mismatch to particular proximate causes and chains of events, problems of one form or another will be inevitable unless the underlying mismatch is addressed.

1.1.6 Subdivided systems

Even if the complexity of the system matches that of its environment at the appropriate scales, there is still the possibility of a complexity mismatch. Consider two pairs of friends—four people total, each of whom can lift 100 pounds—and consider two 200-pound couches that need to be moved. Furthermore, assume that each person is able to coordinate with her friend but not with either of the other two people. Overall then, the system of people has sufficient complexity at the appropriate scales to move both couches since each pair of friends can lift one of the 200-pound couches. However, were one person from each pair of friends to be assigned to each couch, they would not be able to lift the couches because the two people lifting each couch would not belong to the same pair of friends and thus would not be able to coordinate their actions. The problem here is that while the pairs of friends possess enough overall complexity at the right scales to lift the couches, the subdivision within the system of friends is not matched to the natural subdivision within the system of couches. The mismatch in complexity can be seen if we focus our attention on just a single couch: while the couch requires coordinated action at the scale of 200 pounds, the two people lifting it are capable only of two independent actions, each at the scale of 100 pounds.

The way in which academic departments are organized provides a more realistic example of the potential of subdivision mismatch. Academia has multiple levels of subdivision (departments, subfields, etc.) in order to organize knowledge and coordinate people, resulting in a high overall degree of complexity across multiple scales, where scale could refer to either the number of coordinated people or the amount of coordinated knowledge, depending on which aspect of the academic system is under consideration. Similarly, there are multiple

they cannot, from first principles, determine precise quantities such as the amount of heat generated by a phase transition or the temperature at which it occurs (such quantities must be determined empirically).

levels of natural subdivision in the set of problems that academia can potentially address, with each subdivision of problems requiring particular types of coordinated knowledge and effort in order to be solved. Academia's complexity across multiple scales allows it to effectively work on many of these problems. However, there may exist problems that academia, despite having sufficient overall multi-scale complexity, is nonetheless unable to solve because the subdivisions within the problem do not match the subdivisions within academia. The increase in interdisciplinary centers and initiatives over the past few decades suggests the perception of such a mismatch; however, the structure of the academic system as a whole may still hinder progress on problems that do not fall neatly within a discipline or sub-discipline [254, 71, 135, 277, 213, 186].

The above examples provide an illustration of the principle that in order for a system to differentially react to a certain set of behaviors in its environment, not only must the system as a whole have at least as much complexity at all scales as this set of environmental behaviors (as described in section 1.1.5), but also *each subset* of the system must have at least as much complexity at all scales as the environmental behaviors corresponding to that subset. A good rule of thumb for applying this principle is that decisions concerning independent parts or aspects of a system should be able to be made independently, while decisions concerning dependent parts of the system should be made dependently. It follows that the organizations that make such decisions should be subdivided accordingly, so that their subdivisions match the natural divisions in the systems with which they interact.[7]

1.1.7 Hierarchies

A common way in which systems are organized is through hierarchies. In an idealized hierarchy, there are no lateral connections: any decision that involves multiple components of the hierarchy must pass through a common node under whose control these components all (directly or indirectly) lie. The complexity profile of such a hierarchy depends on the rigidity

[7]The subdivisions present in the human brain and the analysis of subdivisions in neural networks more generally [40, Chapters 2.4-2.5] demonstrate how systems that are subdivided so as to match the natural subdivisions in their environments outperform those with more internal connectivity.

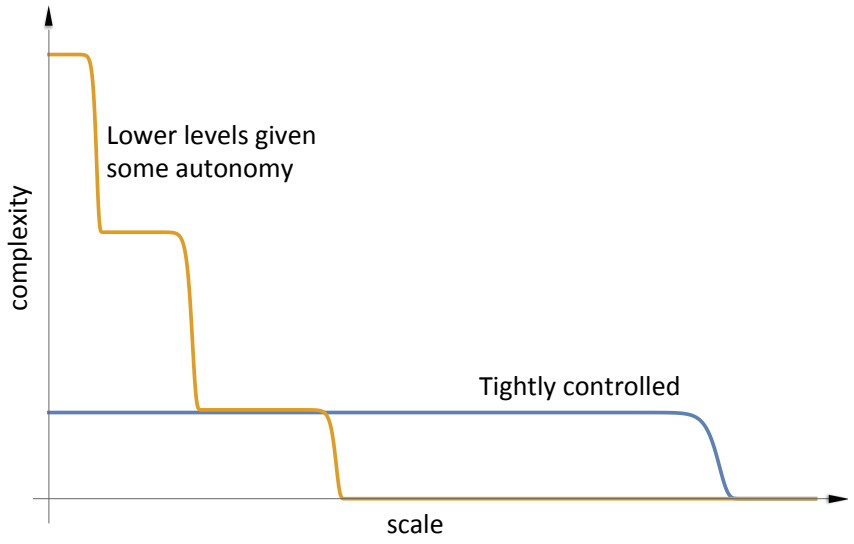


Figure 1-5: Complexity profiles of two hierarchies, each with the same number of people. Here, the scale is the number of coordinated man-hours. In one hierarchy, all decisions, regardless of the scale, are made by a single person, while in the other, different decisions are made at various levels of the hierarchy.

of the control structure (fig. 1-5). At one extreme, every decision, no matter how large or small, is made by those at the top of the hierarchy. This hierarchy has the same amount of complexity across all its scales: namely the complexity of whatever decisions are being made at the top. At the other extreme, there is no communication within the hierarchy, and every individual acts independently. This hierarchy has very little complexity beyond the individual level. Between these two extremes is a typical hierarchy, in which different decisions are made at different levels.

No type of hierarchy is inherently better than any other. For a particular environment, the best hierarchy is one for which the complexity profile matches that of the tasks needed to be performed. A tightly controlled (top-heavy) hierarchy is not well suited to environments in which there is a lot of variation in the systems with which the lower levels of the hierarchy must interact; neither is a very loosely controlled hierarchy well suited to environments that require large-scale coordinated action. For example, centralizing too much power within

the U.S. governance system at the federal (as opposed to the local or state) level would not allow for sufficient smaller-scale complexity to match the variation among locales; too decentralized a governance system would not allow for sufficient larger-scale complexity to engage with problems that require nationally coordinated responses.[8] Assigning decisions to higher levels in hierarchies allows for more efficiency and scale, but such decisions result in less adaptability because when they are incorrect, they affect more of the system and—as larger-scale changes tend to require longer time-scales to enact—are more difficult to roll back.

It is important to distinguish between the complexity of a hierarchy and the complexity of the decisions that the people within the hierarchy are capable of making. For instance, one could design a tightly controlled hierarchy that could take a large number of large-scale actions (i.e. high complexity at its largest scale), but since the decision-making abilities of even the most capable humans are of finite complexity, the individuals at the top may be fundamentally unable to correctly choose from among these actions. This brings us to an important limitation of hierarchies: the complexity of the decisions concerning the largest-scale behaviors of a hierarchy—the behaviors involving the entire organization—is limited by the complexity of the group of people at the top [41]. Thus, a hierarchy will necessarily fail when the complexity of matching its largest-scale behaviors to those of its environment[9] is higher than the complexity of decision-making that is achievable by any individual or committee. The failure of command economies provides a stark example: the allocation

[8]We can consider not just the overall complexity profile of governance systems but also how well the subdivisions in governance systems match those within their territories (section 1.1.6). Metropolitan areas are in some ways more similar to one another than they are to the rural areas of their respective states. So while dividing the U.S. into 50 states provides substantial lower-scale governmental complexity, this complexity is not necessarily well matched to natural urban-rural divides. To the extent that such a mismatch exists, there may be issues currently handled at the state level that would be better handled at the local level, thereby allowing for different policies in urban and rural areas (and likewise, perhaps some of the powers that some argue should be devolved from the federal to the state level should in fact be devolved to the local level).

[9]Note that the complexity of deciding which behaviors of a system should correspond to which behaviors of its environment is generally much greater than the complexity of either the system or the environment alone: for example, if both the system and environment have 10 possible behaviors, the system has enough complexity to match the environment, but properly deciding which behaviors of the system should correspond to which environmental conditions requires correctly choosing one option out of a space of 3,628,800 (10 factorial) possibilities. The space of possible behaviors of a system and its environment may be much smaller than the space of possible decisions concerning the management of the system's actions in its environment.

of resources and labor is too complex a problem for any one person or group of people to understand. Markets allocate resources via a more networked system: decisions regarding how to allocate resources are made without any individual making them, just as decisions are made in the human brain without any neuron making them. (Whether or not these market allocations are desirable depends in part on the way in which the market is structured and regulated.)

We began by considering idealized hierarchies with only vertical connections, but lateral connections provide another mechanism for enabling larger-scale behaviors. For instance, cities can interact with one another (rather than interacting only with their state and national governments) in order to copy good policies and learn from each other's mistakes. Through these sorts of evolutionary processes (described further in section 1.3), large-scale decisions (large-scale because policies may be copied by multiple cities) that are more complex than any individual component can be made. Such lateral connections can exist within a hierarchical framework in which the top of the hierarchy (in this example, the national government) maintains significant control, or they can exist outside of a hierarchical structure, as in the human brain. Furthermore, these lateral connections can vary in strength. Overly strong connections lead to herd-like behaviors with insufficient smaller-scale variation, such as groupthink [187, 330, 259] (no system is exempt from the tradeoff described in section 1.1.4), while overly weak connections result in mostly independent behavior with little coordination.

1.2 Analyzing Complex Systems

The previous section has examined some of the general properties of systems with many components. But how do we study particular systems? How do we analyze data from complex systems, and how do we choose which data to analyze?

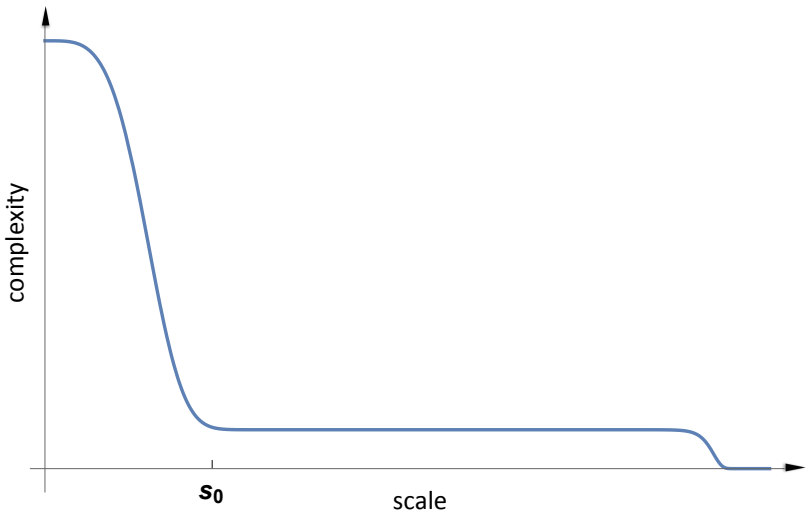


Figure 1-6: A complexity profile of a system with a separation of scales. A separation of scales implies that the behaviors occurring below a certain scale (s_0 in the above figure) are at larger scales mostly independent from one another and that therefore, at these larger scales, only the average effects of the small-scale behaviors are relevant.

1.2.1 How do we understand any system?

In a sense, it is surprising that we can understand any macroscopic system at all, as even a very simple mechanical system has trillions upon trillions of molecules. We are able to understand such systems because they possess a *separation of scales* [50], meaning that the macroscopic behavior we are interested in occurs at a far larger scale than the behavior of the individual molecules, with not much behavior occurring in between these two scales (see fig. 1-6). This separation allows us to treat the macroscopic and microscopic behaviors separately: for mechanical systems, we treat the macroscopic behavior explicitly with Newtonian mechanics, while the microscopic behavior is considered in aggregate using thermodynamics.

More generally, the approach described above is an example of a *mean-field theory* [192], in which the average behaviors of a system's components are explicitly modeled and the deviations of the individual components from this average are treated as statistically independent random fluctuations. This approach works very well for systems such as computers,

cars, airplanes, and buildings, in which the motions of individual molecules are—apart from some mostly uncorrelated fluctuations—well described by the motion of the piece of material to which they belong. Mean-field assumptions are also often employed in analyses of biological, social, and economic systems; these assumptions work well in many cases, but, as we will see, they are not always appropriate for complex systems. It is important, therefore, to determine under what conditions mean-field theory holds.

1.2.2 When mean-field theory breaks down

The systems for which mean-field theory applies exhibit large-scale behaviors that are the average of the behaviors of their components. They must possess a separation of scales, which arises when the statistical fluctuations of their components are sufficiently independent from one another above a certain scale. Mean-field theory may hold even in the presence of strong interactions, so long as the effect of those strong interactions can be captured by the average behavior of the system—that is, so long as each component of the system can be modeled as if it were interacting with the average (i.e. mean field) of the system. For example, the large-scale motion of solids is well described by mean-field theory, even though the molecules in a solid interact with one another quite strongly, because the main effect of these interactions is to keep each molecule at a certain distance and orientation from the average location (center of mass) of the solid. Likewise, under some (but certainly not all) conditions, economic markets can be effectively described by modeling each market actor as interacting with the aggregate forces of supply and demand rather than with other individual market actors.

However, when there are sufficiently strong correlations between the components of the system, i.e. when the interactions between a component of the system and a specific set of other components (as opposed to its general interaction with the rest of the system) cannot be neglected, mean-field theory will break down.[10] These systems will instead exhibit large-scale behaviors that arise not solely from the properties of individual components but also

[10]Incidentally, the failure of mean-field theory to describe certain physical phase transitions led physicists to develop a new, multi-scale approach (the renormalization group), a foundation of much of complex systems science.

from the relationships between components. For example, while the behavior of a muscle can be roughly understood from the behavior of an individual muscle cell, the behavior of the human brain is fundamentally different from that of individual neurons, because cognitive behaviors are determined largely by variations in the synapses *between* neurons. Similarly, the complex ecological behaviors of a forest cannot be determined by the behaviors of its constituent organisms in isolation.

Because their small-scale random occurrences are not statistically independent, complex systems often exhibit large-scale fluctuations not predicted by mean-field theory, such as forest fires, viral content on social media, and crashes in economic markets. Sometimes, these large-scale fluctuations are adaptive: they enable a system to collectively respond to small inputs [252]. (For instance, humans respond strongly to minor disturbances in the density of air, such as the sound of their own names.) However, these large-scale fluctuations sometimes pose systemic risks.

1.2.3 Fat-tailed distributions and systemic risk

When the components of a system are independent from one another above a certain scale, then at much larger scales, the magnitude of the fluctuations of the system follow a normal distribution (bell curve),[11] for which the mean and standard deviation are well-defined and for which events many standard deviations above the mean are astronomically improbable. Interdependencies, however, can lead to a distribution of fluctuations in which the probability of an extreme event, while still small, is not astronomically so. Such distributions are characterized as *fat-tailed*—see fig. 1-7. For example, while human height follows a thin-tailed distribution, with no record of anyone over twice as tall as the average human, human wealth—due to the complex economic interactions between individuals—follows a fat-tailed distribution, with multiple individuals deviating from the average by factors of more than one million [110].

[11] This follows from the central limit theorem.

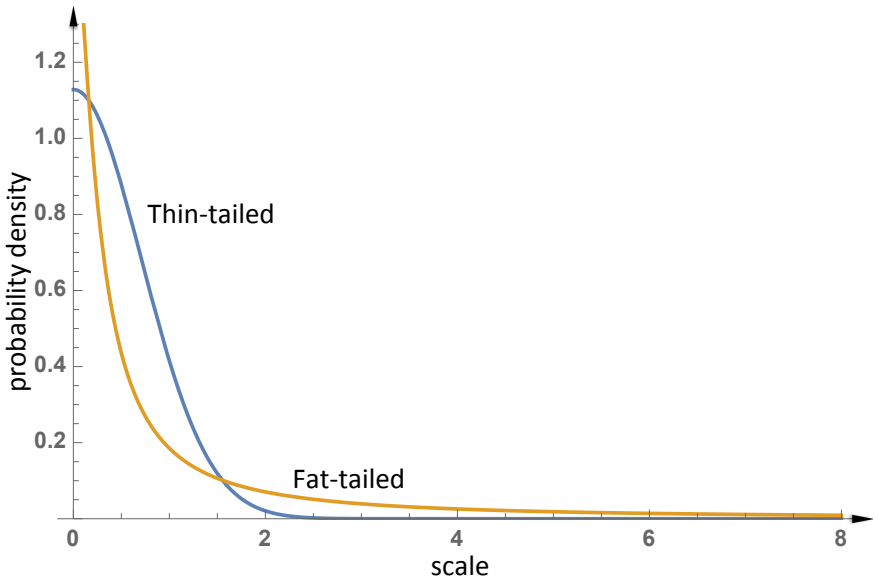


Figure 1-7: A normal distribution (thin-tailed) and a distribution with a power-law decay (fat-tailed). The fat-tailed distribution may appear more stable, due to the lower probability of small-scale fluctuations and the fact that samples from the distribution may not contain any extreme events. However, sooner or later, a fat-tailed distribution will produce an extreme event, while one could wait thousands of lifetimes of the universe before a normal distribution produces a similarly extreme event. Note that the axes of this graph are truncated; the illustrated fat-tailed distribution can, with small but non-negligible probability (0.04%), produce events with a scale of one million or more.

One danger of interdependencies is that they may make systems appear more stable in the short term by reducing the extent of small-scale fluctuations, while actually increasing the probability of catastrophic failure [76, 340, 168, 37]. This danger is compounded by the fact that when underlying probability distributions have fat tails (a situation made more likely by interdependencies), standard statistical methods often break down, leading to potentially severe underestimates of the probabilities of extreme events [320].[12] As a thought experiment, imagine 100 ladders, each with a 1/10 probability of falling. If the ladders are independent from one another, the probability that all of them fall is astronomically low (literally so: there is about a 10^{20} times higher chance of randomly selecting a particular atom out of all of the atoms in the known universe). If we tie all the ladders together, we will have made them safer, in the sense that the probability of any individual ladder falling will be much smaller, but we will have also created a non-negligible chance that all of the ladders might fall down together. Other examples include the interconnectedness of our financial systems resulting in the possibility of global market crashes [239, 165, 161, 134, 292, 305] and the interconnectedness of travel routes increasing the probability of pandemics such as the Spanish flu and COVID-19 [275, 297]. When such crises do occur, they are often attributed to proximate causes or chains of events, and measures are then implemented to ensure that those particular chains of events will not occur again. But unless the underlying systemic instabilities are addressed, another crisis is bound to happen sooner or later, even if its precise form cannot be predicted.

1.2.4 Understanding complex systems

Because it is usually easier to collect data regarding components of a system than it is to collect data regarding interactions between components, studies often fail to capture the information relevant to complex systems, since complex large-scale behaviors critically depend on such interactions. Furthermore, as discussed in section 1.2.3, data analysis can

[12] For instance, the mean and variance of a fat-tailed distribution may not be well-defined, and even if they are, they may not be able to be reliably estimated from a finite sample due to the likelihood of a single data point or lack thereof substantially skewing the estimate.

severely underestimate the probability of extreme events (tail risk). Finally, analyses often (implicitly) assume linearity, i.e. they assume that the total impact of a set of factors is equal to the sum of the impacts of each individual factor, an assumption that often breaks down for complex systems, which may possess feedback loops, abrupt transitions (tipping points), and other highly nonlinear behaviors [226, 170, 288, 289, 169, 282, 351, 299].

How can we understand the systems for which these standard approaches do not apply? Our understanding of *all* systems with many components depends on *universality* [193], i.e. the existence of large-scale behaviors that do not depend on the microscopic details. The standard approaches are predicated on the assumption of sufficient independence between components, which allows large-scale behaviors to be determined without a full accounting of the system's details via mean-field theory.[13] But mean-field theory is just one example of universality.

Sound is another example: all materials, regardless of their composition, allow for the propagation of sound waves. Sound behaves so similarly in all materials because at the length scales relevant to sound waves, which are far larger than the sizes of individual atoms and molecules, the effect of the microscopic parameters is merely to set the speed of the sound.[14] Note that sound waves cannot be understood as a property of the average behavior—in this case, average density—of a material, since it is precisely the systematic correlations in the deviations from that average that give rise to sound. Nor is sound best understood by focusing on the small-scale details of atomic motion: scientists understood sound even before they learned what atoms are. The key to understanding sound waves is to recognize that they have a multi-scale structure—with larger-scale fluctuations corresponding to lower frequencies and smaller-scale fluctuations corresponding to higher frequencies—and to model them accordingly.

[13]Formally, the statistical fluctuations of the components must, above a certain scale, be sufficiently independent so as to satisfy the assumptions of the central limit theorem. The central limit theorem is the manifestation of universality that explains the ubiquity of normal distributions.

[14]For Quantum Electrodynamics (the theory of how light and electrons interact), we *still do not know* the microscopic details. Yet we can nonetheless make predictions accurate to ten decimal places because, as can be shown with renormalization group theory, the only effect of these microscopic details at the scales at which we can make measurements is to set the electron mass and charge—quantities that, like the speed of sound in any particular material, can be measured (but not predicted).

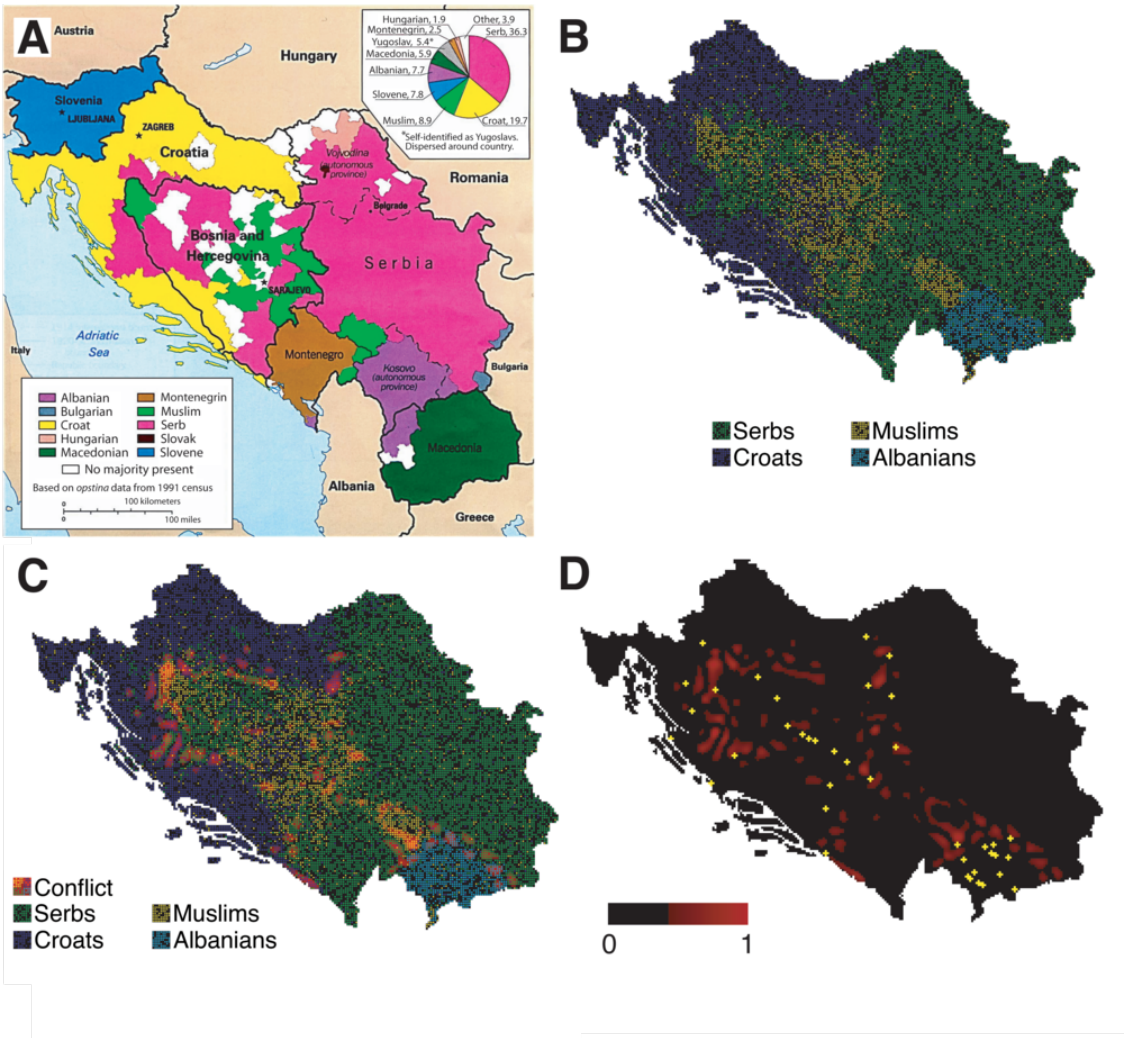


Figure 1-8: A figure from Lim *et al.*'s paper on ethnic violence [218]. The sites where their model predicts a potential for ethnic violence are shown in red in panels C and D, with confirmed reports of ethnic violence depicted by the yellow dots in panel D.

Lim *et al.* apply this approach to studying ethnic violence [218]. They built a predictive model to analyze where ethnic violence has the potential to occur and applied their model to India and to what was Yugoslavia. Ethnic violence has many causes, but rather than focusing on specific, culturally dependent mechanisms or on the average properties of regions, such as demographic or economic statistics, the authors instead considered the multi-scale patterns in how ethnic groups were geographically distributed (fig. 1-8). They found that ethnic violence did not occur when the ethnic groups were either well mixed or well separated but rather occurred only when ethnic groups separated into geographic patches,[15] with the violence most likely to occur for geographic patches of a particular size.[16] Although not explicitly included in the analysis, specific details of a region are relevant insofar as they are either a cause or an effect (or both) of the patch size.[17]

Understanding all of the details of any complex system is impossible, just as it is for most systems with a separation of scales; there is just too much complexity at the smallest scale. But unlike the behaviors of systems with a separation of scales, the important large-scale behaviors of complex systems are not simply the average of their small-scale behaviors. The interdependencies at multiple scales can make it difficult or impossible to precisely understand how small-scale behaviors give rise to larger-scale ones, but even for complex systems, there is much less complexity at the larger scales than there is at the smaller scales. Thus, there will always be large-scale behaviors that do not depend on most of the system's details (see fig. 1-9). The key to analyzing these behaviors is to find the appropriate mathematical (or conceptual) description—i.e. to identify variables that describe the relevant space of possible (large-scale) behaviors—which for complex systems is not a simple average nor a full account of all the details. For additional examples of this multi-scale approach, see ref. [50].

[15]This separation falls into the same universality class as the separation of oil and water.

[16]This analysis implies that ethnic violence can be prevented by the use of well-placed political boundaries, as in Switzerland [281].

[17]For instance, animosity between two ethnic groups, though not explicitly considered, may be a cause as well as a consequence of the geographic segregation [290].

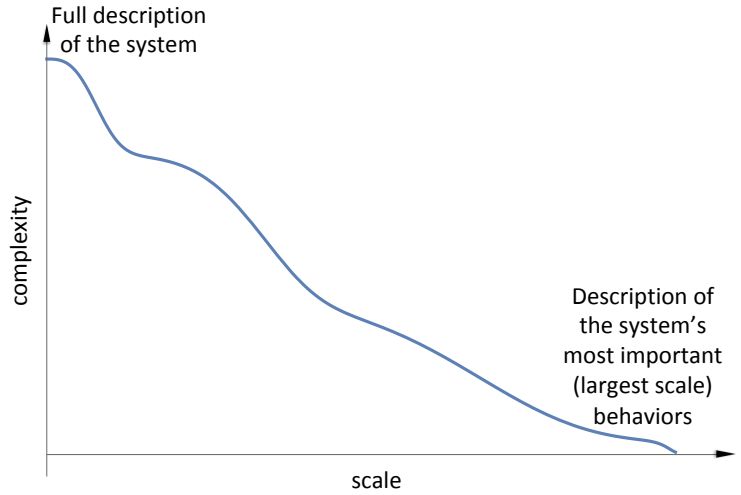


Figure 1-9: A representative complexity profile of a complex system. Understanding all the details (i.e. all of the small-scale behaviors) is impossible and unnecessary; the most important information is contained in the large-scale behaviors. However, for systems for which mean-field theory does not apply, characterizing these behaviors will involve more than a simple average.

1.3 Complex Systems and Uncertainty

Although the principles discussed throughout sections 1.1 and 1.2 help us recognize the fundamental properties and limitations of systems, our understanding of most complex systems will inevitably be imperfect. And regardless of how well-considered a plan is, a truly complex system will present elements that were not considered ahead of time.[18] Given the absence of perfect knowledge, how can the success of systems we design or are part of be assured? While the success of many systems rests on the assumption that good decisions will be made, some systems do not depend on individual understanding and can perform well in spite of the fallibility of decision-makers (whether due to corruption, subconscious bias, or the fundamental

[18]It should also be noted that in a functional system with a high degree of complexity, the potential positive impact of a change is generally much smaller than its potential negative impact. For example, a change to the wiring in a computer is unlikely to dramatically improve the computer's performance, but it could cause the computer to crash. Airplanes are another example. This phenomenon is a consequence of the fact that, by definition, a high degree of complexity implies that there are many system configurations that will not work for every one configuration that will.

limitations of human minds). The study of complex systems approaches this observation scientifically by (implicitly or explicitly) considering the decision-makers themselves as part of the system and of limited complexity/decision-making ability. The question thus becomes: how do we design systems that exceed the complexity of the decision-makers within them?

1.3.1 Evolutionary processes

While uncertainty makes most systems weaker, some systems benefit from uncertainty and variability [331, 241, 154, 319]. The common characteristic of these systems is their embodiment of some sort of evolutionary process, i.e. a process in which successful changes are copied (and further modified) while unsuccessful changes are not. The classic evolutionary processes are biological: due to variability introduced by random mutations, organisms with the complexity and scale of humans evolved from single-celled organisms. Furthermore, humans themselves have the property of benefiting from exposure to random shocks (provided the shocks are not too strong). Immune system performance is improved by early exposure to non-lethal pathogens [256, 312]; muscles and bones are strengthened by micro-tears and micro-fractures, respectively; we learn by exposure to new information and problem-solving; and our psychologies are strengthened by exposure to adversity, provided the adversity is not too severe [293, 227].

Competitive market economies provide another example of how systems can thrive on uncertainty. Due to our ignorance of which will succeed, many potential innovations and businesses must be created and improved upon in parallel, the successful ones expanding and the unsuccessful ones failing. The successful among these can then be improved upon in the same manner—with many approaches being applied at once—and so on. (However, without effectively regulated multi-scale cooperative frameworks—see section 1.3.2—large-scale parts of the economic system may optimize for the wrong goals, settling into harmful societal equilibria [257, 228].)

Likewise, the internal processes of large organizations may follow an evolutionary pattern in which small parts of the organization can fail and thus be improved upon; without such

flexibility, the entire organization may fail at once in the face of a changing internal or external environment. In some cases the failure of the entire organization makes room for more effective organizations to take its place (assuming the economy is sufficiently decentralized and competitive so that the organization in question is not "too big to fail"). The collapse of government is generally not one of those cases, however [142], so it is especially important that governance systems possess the flexibility to internally benefit from randomness and uncertainty. Perhaps counterintuitively, not allowing small failures to occur may weaken systems in the long run by halting evolutionary processes and by creating interdependencies that lead to systemic risk (section 1.2.3).

In order to thrive in uncertainty and exceed the complexity of individual decision-making, systems can incorporate evolutionary processes so that they, even if very limited at first, will naturally improve over time. The first step is to allow for enough variation in the system, so that the system can explore the space of possibilities. Since a large amount of variation means a lot of complexity and complexity trades off with scale (section 1.1.4), such variation must occur at smaller scales (in both space and time). For example, in the case of governance, enabling each city to experiment independently allows for many plans to be tried out in parallel and to be iterated upon. The opposite strategy would be to enact one national plan, the effects of which will not be able to be comparatively evaluated.

The second step is to allow for a means of communication between various parts of the system so that successful choices are adopted elsewhere and built upon (e.g. cities copying the successful practices of other cities). Plans will always have unintended consequences; the key is to allow unintended consequences to work for rather than against the system as a whole. The desire for direct control must often be relinquished in order to allow complexity to autonomously increase over time.[19]

[19]Systems can explicitly design only systems of lesser complexity since an explicit design is itself a behavior of the first system. However, systems that evolve over time can become more complex than their designers.

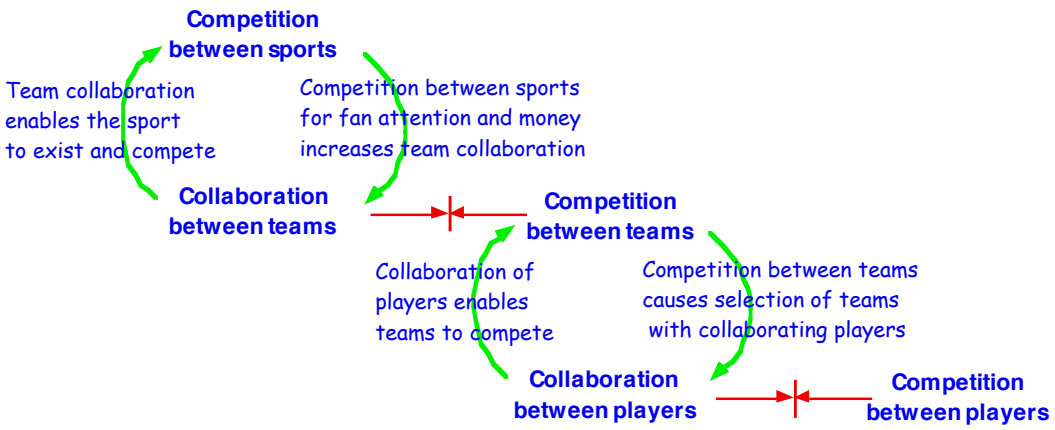


Figure 1-10: An illustration from ref. [43, Chapter 7] showing the interplay between cooperation and competition in the context of sports teams and leagues.

1.3.2 Multi-scale evolutionary processes

Successful evolutionary processes generally do not consist of unbridled competition but rather contain both competition and cooperation, each occurring at multiple scales [47]. For example, cells cooperate within multicellular organisms in order to more effectively compete with other organisms, and organisms cooperate both within and between species in order to more effectively compete against other species. Competition at larger scales naturally breeds cooperation at smaller scales because in order for a group to effectively compete against another group (large-scale competition), there must be cooperation within the group. Cooperation can also breed competition since sometimes the best way for the group to achieve its shared goals is to facilitate some healthy competition among its subgroups. Those subgroups must foster cooperation within themselves in order to effectively compete with each other, and they too may be able to increase the effectiveness of their internal cooperation by introducing some healthy competition among their members (section 1.3.1 provides an example). If these members are themselves groups, the process of competition begetting cooperation that begets more competition can continue to even smaller scales. This process can work in reverse as well: in order for individuals to compete more effectively, they may cooperate with each other to form groups, which in turn may cooperate to form even larger groups, and so

on. Thus, a complex network of cooperation and competition among groups of various sizes (scales) can naturally evolve.

In order to promote effective group cooperation, competition must be properly structured. A soccer team in which the players compete with their own team members to score goals will not be effective, but one in which the players compete for the title of the most fit may be. The framework in which competition occurs must be structured so that the competitors are incentivized to take actions that are net good for the group; otherwise a kind of tragedy-of-the-commons situation occurs. The potential for competition to go awry highlights the importance of having a multi-scale structure with competition occurring on multiple levels, rather than having everyone in the system compete with everyone else. With the multi-scale structure, groups with unhealthy evolutionary dynamics are selected against, while groups with a healthy mix of competition and cooperation that benefits the entire group are selected for.[20] Market economic systems are successful not because free markets produce optimal outcomes (real-world markets often sharply deviate from the assumptions of free-market models, and externalities abound) but rather because, at their best, appropriately regulated market systems allow for multi-scale evolutionary processes to naturally arise, resulting in innovations and complexity far beyond what anyone could have imagined, let alone designed.

1.4 Further reading

Complex systems science, also known as complexity science, contains many subfields. One starting point for exploring complex systems more broadly is this clickable map [83] of complex systems science and related fields. Encyclopedias [335, 246] and textbooks [302, 40, 251, 249, 286, 327] provide a range of perspectives. In addition to the topics and references discussed throughout this introduction, we provide a selection among the many works

[20]There is evidence that the geographic nature of evolution—in which organisms evolve in somewhat separated environments and mean-field theory does not apply—has resulted in precisely this multi-scale structure and has therefore allowed for the evolution of genuine (e.g. not reciprocal) altruistic behavior [350, 352].

applying complex systems science to social systems and policy [78, 214, 38, 153, 325, 238, 162, 26, 79, 36, 222, 148, 49, 126, 311] and management [346, 307, 176, 163]. Complex systems science includes, among others, the fields of system dynamics [309], evolutionary dynamics [194, 303, 24], network science [52], fractals and scaling [115, 236, 28, 73], urban science [58], pattern formation [328, 291], econophysics [237], and nonlinear dynamics and chaos [310, 152]. Book series on complex systems topics include the *Santa Fe Institute Series* and *Unifying Themes in Complex Systems*.

1.5 Summary

Systems with many components often exhibit emergent large-scale behaviors that cannot be directly inferred from the behaviors of their components. However, an early insight of statistical physics is that in spite of the impossibility of describing the details of trillions of molecules, the macroscopic properties of the molecules can be well understood by analyzing their space of possible behaviors, rather than their specific configurations and motions. While many macroscopic properties can be described in terms of the average behaviors of the molecules, the macroscopic properties of certain physical phenomena, such as phase transitions, cannot be understood by averaging over system components; accordingly, physicists were forced to develop new, multi-scale methods. Likewise, while standard statistical methods—which infer the average properties of a system's many components—can successfully model some biological and social systems, they fail for others, sometimes spectacularly so.

Taking a systemic view by considering the space of possible behaviors can yield insights that cannot be gleaned by considering only the proximate causes and effects of particular problems or crises. A system's complexity—which depends on its number of distinct potential behaviors (i.e. on the space of possibilities)—is a starting point from which to get a handle on its large-scale properties, in the same way that entropy is the starting point for statistical physics. Because the number of distinct behaviors of a system depends on the level of detail (behaviors that appear the same at lower resolution may be distinct at

higher resolution), complexity depends on scale. Interdependencies between components reduce complexity at smaller scales by restricting the freedom of individual components while creating complexity at larger scales by enabling behaviors that involve multiple components working together. Thus, for systems that consist of the same components, there is a fundamental tradeoff between the number of behaviors at smaller and larger scales. This tradeoff among scales is related to the tradeoff between a system's adaptability, which depends on the variety of different responses it has to internal and external disturbances, and its efficiency, which depends on its operating scale. There is no ideal scale at which a system should possess complexity; rather, the most effective systems are those that at each scale match the complexity of their environments.

When analyzing data or creating organizational structures, standard methods fail when they underestimate the importance of interdependencies and the complexity that arises from these interdependencies. To some extent, these problems can be mitigated by matching the data analysis or organizational structure to natural divisions within the system of interest. Since complex systems are those for which behaviors occur over multiple scales, successful organizations and analyses for complex systems must also be multi-scale in nature. However, even when armed with all the proper information and tools, human understanding of most complex systems will inevitably fall short, with unpredictability being the best prediction. To confront this reality, we must design systems that are robust to the ignorance of their designers and that, like evolution, are strengthened rather than weakened by unpredictability. Such systems are flexible with multiple processes occurring in parallel; these processes may compete with one another within a multi-scale cooperative framework such that effective practices are replicated. Only these systems—that grow in complexity over time from trial and error and the input of many—exhibit the necessary complexity to solve problems that exceed the limits of human comprehension.

Chapter 2

A Formal Definition of Scale-dependent Complexity and the Multi-scale Law of Requisite Variety

2.1 Introduction

A problem that has plagued the field of complex systems science is the need for a general definition of complexity. Scale-dependent complexity [61, 41, 45, 149, 245, 25, 51, 12, 298] offers a promising path: rather than attempt to describe the complexity of a system with a single number, it recognizes that the difficulty in describing a system (complexity) depends on the level of detail of the description (scale). The paradox as to whether a human or a gas containing the same atoms of that human is more complex is resolved: the gas has more complexity at the smallest scale but the human has more complexity at larger scales (section 1.1.3).

Another key concept in complex systems science is Ashby's law of requisite variety [21], which states that a system must have at least as many behaviors as the number of environmental behaviors to which it must differentially respond. In other words, a system should have more complexity than its environment, if the environment is defined to include only the

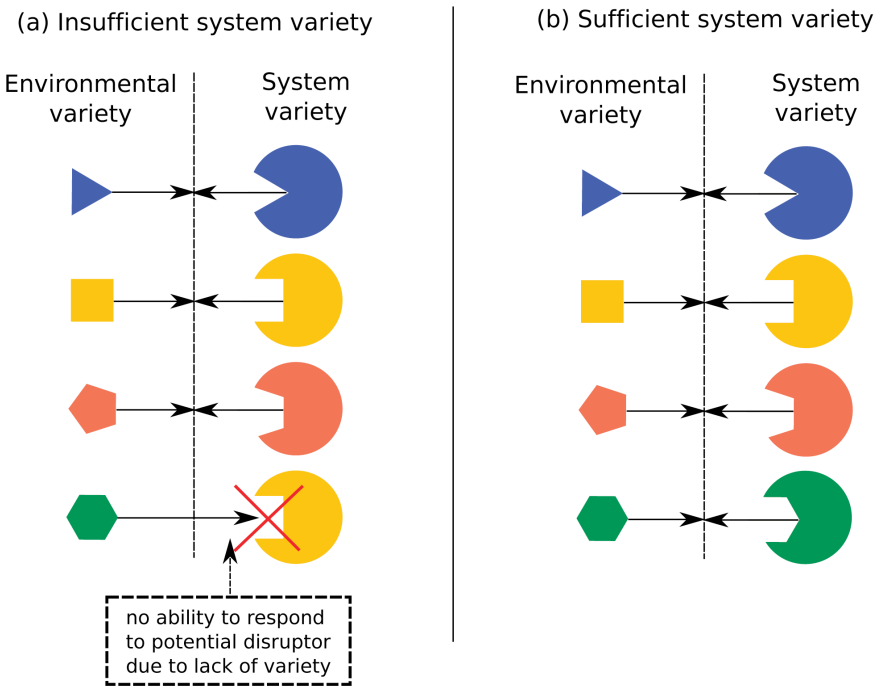


Figure 2-1: An illustration of Ashby's law of requisite variety. (a) Because the system has fewer states (i.e. lower complexity) than its environment, it is impossible for the system to have a distinct response to each of the four environmental states. (b) Here, the system is able to have a distinct response to each environmental state; a necessary (but not sufficient) condition for this matching is that the system's complexity equals or exceeds its environment's. Image source: ref. [255].

relevant behaviors to which the system must respond (see fig. 2-1). Since, however, complexity depends on scale, it would be valuable to be able to evaluate a system's ability to respond to its environment at multiple scales at once, rather than having to redefine the space of possible behaviors for each scale. In other words, it would be valuable for Ashby's law to apply to the complexity profile at each scale. For a pedagogical introduction to complexity profiles and the law of requisite variety, please see section 1.1; the purpose of this chapter is to put a mathematical formalism behind these concepts, and thus we assume the reader is already familiar with their conceptual utility.

For the current formal definition of a complexity profile, it has been proven that a multi-scale law of requisite variety applies for systems and environments that are *block-independent*,

where a system of components is block-independent if the components can be partitioned such that components within the same partition (block) have identical behavior, with all of the blocks being mutually independent from one another [46]. However, as we will see, Ashby's law does not apply to this definition of a complexity profile more generally. Given the desirability and usefulness of a complexity profile satisfying Ashby's law at each scale (a property that has been implicitly used in many analyses [41, 43, 44, 42, 255, 298]), we therefore seek a formal definition of the complexity profile that reflects this property. The one other property that we wish for a complexity profile to have is a sum-rule, that the area under the complexity profile does not depend on interdependencies between components but rather only the individual components' behaviors. Such a constraint leads to a very useful tradeoff formalization of the tradeoff between complexities at various scales.

In order to define a complexity profile for which Ashby's law holds, we have to define what it means for a system to effectively match its environment. Thus, in section 2.2, we explore the law of requisite variety and define a multi-component version of it. In section 2.3, we formally define what constitutes a complexity profile and what criterion must be satisfied for it to capture the multi-scale law of requisite variety and the tradeoff between complexities at various scales. In section 2.4, we define a class of complexity profiles that satisfy such criteria. Such a class does not provide a single complexity profile, fitting with the idea that which behaviors that are large-scale or coarse-grained are to some extent a subjective choice. That such a subjective choice exists does not mean that there are not, in many cases, choices that are far more useful than others. But it does mean that there is not *a priori* a single best way to aggregate or coarse-grain small-scale details into relevant large-scale behaviors.

2.2 Generalizing Ashby's Law to Multiple Components

Ashby's law claims that to effectively regulate an environment, the system must have a degree of freedom or behavior for each distinct environmental behavior. In other words, there cannot be two environmental states for a given system state. It then follows (by the pigeon-hole principle) that the number of behaviors of the system must be greater than

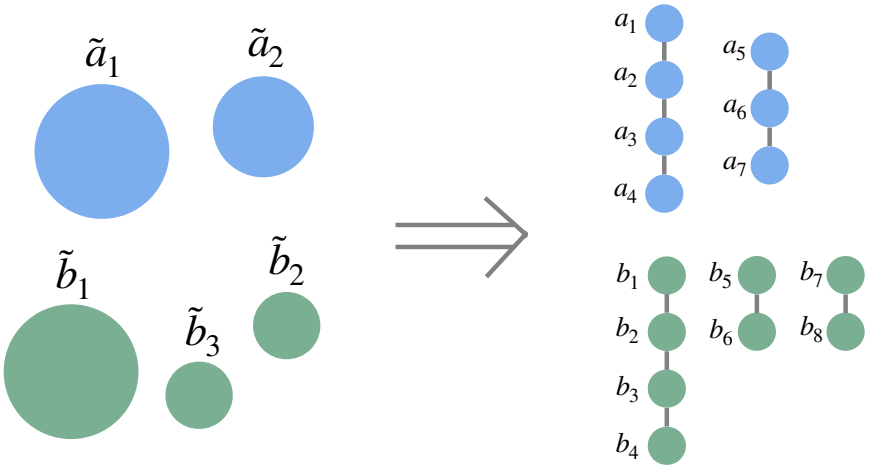


Figure 2-2: Accounting for components of various sizes. All systems can be described by components that are all of the same size. For instance, two systems $\{\tilde{a}_1, \tilde{a}_2\}$ and $\{\tilde{b}_1, \tilde{b}_2, \tilde{b}_3\}$ that contain components of sizes $2s$, $1.5s$, and $1s$ (where the units of s depend on the notion of size being used) can be reformulated in terms of components $\{a_1, ..., a_7\}$ and $\{b_1, ..., b_8\}$ that all have size $0.5s$, where $a_1 = a_2 = a_3 = a_4 = \tilde{a}_1$, $a_5 = a_6 = a_7 = \tilde{a}_2$, $b_1 = b_2 = b_3 = b_4 = \tilde{b}_1$, $b_5 = b_6 = \tilde{b}_2$, and $b_7 = b_8 = \tilde{b}_3$.

or equal to the number of behaviors of the environment. Formally, for a system X and environment Y, $H(Y|X) = 0$, from which it follows that $H(X) \geq H(Y)$, i.e. the complexity of the environment can not exceed that of the system. Throughout this chapter, H can be any generalized information measure [12], of which Shannon entropy is one example. It is important to note that in this formulation, the environment Y is *defined* to be the set of states that require distinct behaviors of the system. Two environmental states that do not require different system behaviors should be represented by a single state of Y.

In order to consider multi-scale behavior, let us describe the system X as consisting of N components such that $X = \{x_1, ..., x_N\}$.

Definition 1. *A* system X *of size* $N = |X|$ *is defined as a set of* N *random variables. These random variables are referred to as* components *of the system.*

Remark 1. *In this formulation, all components of one or more systems are treated equivalently; while this may seem like a limitation, any system or systems can be described in this*

way to arbitrary precision: components of different sizes can be accounted for by defining a new set of components whose size is the greatest common factor of the sizes of the original components (irrational relative sizes—for which no greatest common factor exists—can be approximated to arbitrary precision by rational relative sizes). If the new components are all of size l, each original component $\tilde{x}_i$ of size l_i can then be replaced with l_i/l new components $x_{j+1}, x_{j+2}...x_{j+s_i/s}$ that are all identical to each other, i.e. $x_{j+1} = x_{j+2} = ... = x_{j+s/s_i}$ (see e.g. fig. 2-2).

The assumption that the system X must have at least one distinct response for each environmental state Y (i.e. $H(Y|X) = 0$) is generalized as follows: an "environmental component" y_i is defined for each system component $x_i \in X$, such that each y_i is a random variable containing the environmental states that require a distinct response from the system component x_i. Then, for the system to effectively interact with its environment, $H(y_i|x_i) = 0$ for each i, i.e. there cannot be two environmental component states for a given state in the corresponding system component. This formulation allows for constraints among the environmental components to induce constraints among the system components. Letting $Y = \{y_1, ..., y_N\}$, we see that $H(y_i|x_i) = 0$ implies $H(Y|X) = 0$ and, thus $H(y_i|x_i) = 0$ is a stronger condition: not only must the system match the environment overall, but this matching must be properly organized.

Although it may not seem so at first, any interaction between a system and its environment can be formulated as above: if we start with a more general formulation in which each system component x_i interacts with environmental components $\tilde{y}_{i_1}, \tilde{y}_{i_2}, ..., \tilde{y}_{i_{n_i}}$, which allows for each system component to interact with multiple environmental components and vice versa, then we can simply redefine the environmental components such that each x_i is associated with the random variable $y_i \equiv (\tilde{y}_{i_1}, \tilde{y}_{i_2}, ..., \tilde{y}_{i_{n_i}})$ (see fig. 2-3 for an example).

Definition 2. *An* environment (Y, f) *for system X is a system Y together with a bijection $f : Y \to X$.*

Definition 3. *A system X* matches *its environment (Y, f) iff $H(y|f(y)) = 0$ for all $y \in Y$.*

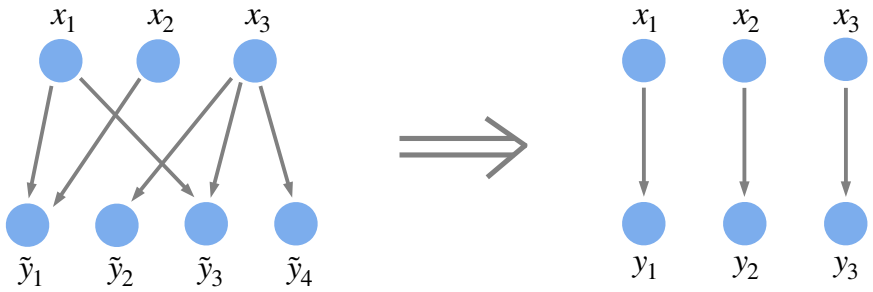


Figure 2-3: Defining the environmental components. Regardless of the interactions between a system and its environment, environmental components can always be defined such that they have a one-to-one relationship with system components. For instance, suppose that for the system $\{x_1, x_2, x_3\}$ to effectively interact with its environment $\{\tilde{y}_1, \tilde{y}_2, \tilde{y}_3, \tilde{y}_4\}$, x_1 must have a distinct response for each possible state of $\tilde{y}_1$ and $\tilde{y}_3$, x_2 must have a distinct response for each possible state of $\tilde{y}_1$, and x_3 must have a distinct response for each possible state of $\tilde{y}_2$, $\tilde{y}_3$, and $\tilde{y}_4$. Environmental components $y_1 \equiv (\tilde{y}_1, \tilde{y}_3)$, $y_2 \equiv \tilde{y}_1$, and $y_3 \equiv (\tilde{y}_2, \tilde{y}_3, \tilde{y}_4)$ can then be defined such that x_1 must react only to y_1, x_2 only to y_2, and x_3 only to y_3.

Example 1. *Consider a system of two thermostats x_1 and x_2 for two rooms that can be either on or off. The environment can be described by two variables y_1 and y_2 that represent whether or not room 1 or room 2, respectively, should be heated (the bijection f mapping y_1 to x_1 and y_2 to x_2). In order for the system to match the environment, it must be that $H(y_1|x_1) = H(y_2|x_2) = 0$. Thus if the need for each room to be heated is independent of that of the other room, the thermostats must be able to operate independently of one another; likewise if the two rooms' need for heat are correlated, the thermostats must also be correlated.*[1]

Example 2. *Consider a system X in which each $x_i \in X$ represents the some aspect of policy (e.g. educational policy) being applied in region i of a given country (the regions could, for instance, be towns/cities). The environment (Y, f) could be defined by the random variables y_i (where $f(x_i) = y_i$), such that each y_i corresponds to conditions in region i that require a distinct policy in order for the region to be effectively governed. If the y_i vary*

[1]Note that this formalism has nothing to say regarding causation. It may be that the objective is to heat both rooms at the same time or not at all, in which case the thermostats themselves must be connected. Or it may be that it just so happens that the two rooms get cold at the same time, in which case two disconnected thermostats may nonetheless exhibit correlated behavior.

independently of one another, while the x_i cannot, then the system X will not be able to match its environment (Y, f) (e.g. an education policy that is homogeneously varied entirely at the national level will not be able to effectively interact with locales if each locale has specific educational needs). Conversely, if there are correlations among the y_i that are lacking in the x_i, the system will also be unable to match its environment (e.g. it would be ineffective for each city to independently set its own policy with respect to international trade or with respect to regulating a national corporation that spans many cities).

Note that the possible states of a system or the probabilities assigned to these states cannot be defined without specifying the environment with which the system is interacting, for the same system may behave differently in different environments. (Alternatively, each individual environment need not be treated separately, since the set of possible different situations that the system may find itself in can itself be considered to be the environment, with the system then being required to differentially respond to—and within—each situation.) In any event, definition of a system matching its environment is ignorant of the mechanism by which the system and the environment are related: a system matching its environment is merely a descriptive statement.

Example 3. *Returning to the example of the two thermostats, if the system (the thermostats) and the environment (the rooms) are connected so that the state of thermostat i depends directly on the state of room i, then the thermostat states will have precisely as much correlation as the room states do. The thermostat states will be independent random variables if and only if the room states are.*

Example 4. *Consider a system of two people in communication with one another. Such a system is versatile in that the individuals could, depending on the circumstances, chose to act either independently or dependently. However, it should be noted that the ability for the two people to choose whether or not they act independently is itself a dependent behavior. Furthermore, seemingly independent behaviors of this system may in fact be dependent if the behaviors of one individual are correlated with the brain state of the other.*

With definition 3, we have a characterization of Ashby's law that takes into account the multi-scale structure of a system and its connection with its environment. The goal is then to understand how properties of the environment constraint the corresponding properties of the system. If an environment has a certain property and it is known that the system matches the environment, what must be true about the system? For the single-scale case of Ashby's law, the system must have at least as much information as the environment. The complexity profile, described below, generalizes this property to multiple scales. In particular, it allows us to formulate the multi-scale law of requisite variety: "in order for a system to match its environment, it must have at least as much complexity as its environment *at every scale*."

2.3 Defining a Complexity Profile

The basic version of Ashby's law states that for a system X to match its environment Y, the overall complexity of X must be greater than or equal to the overall complexity of Y. But, as argued in section 1.1.3, it does not make sense to speak of complexity as a single number but rather the complexity of a system must depend on its scale. Thus, we wish to generalize the notion of a complexity profile such that the complexity of a system and its environment can be compared at multiple scales.

Definition 4. *A complexity profile $C_X(n)$ of a system X assigns a particular amount of information to the system at each scale $n \in \mathbb{Z}^+$. For $n > |X|$, we define $C_X(n) = 0$. If we wish to consider each component of the system to be of size l, we can define a continuous version of the complexity profile (see section 2.6.2 for more detail):*

$$\tilde{C}_X(s) = C_X(\lceil s/l \rceil) \tag{2.1}$$

We wish for a complexity profile to have two additional properties: it should (1) manifest the *multi-scale law of requisite variety* and (2) obey the *sum rule*. Each property is defined below, with applications/examples given in sections 1.1.5 and 1.1.4, respectively.

Definition 5. *A complexity profile manifests the multi-scale law of requisite variety if, for any two systems X and Y, X matching Y (per definition 3) implies that $C_X(n) \geq C_Y(n)$ for all n.*

Definition 6. *A complexity profile obeys the sum rule if for any system X, $\sum_{n=1}^{\infty} C_X(n) = \sum_{x \in X} H(x)$.*

The multi-scale law of requisite variety is important because it allows for the interpretation that a necessary (but not sufficient) condition for a system to effectively interact with its environment must be that it has at least as much complexity as the environment at every scale. The sum rule is important because it captures the intuition that for a system composed of components with the same individual behaviors, there is a tradeoff among the complexities of the system at various scales.

Note that examining measures of multi-scale complexity can never prove that a system matches its environment—just as in the single-scale case, a system having more complexity than its environment by no means guarantees that every system state corresponds to a single environmental state. But examining multi-scale measures of information *can* prove the impossibility of compatibility. The goal then, in formulating multi-scale measures, is to create more instances in which the impossibility of compatibility can be shown. Using this multi-scale formalism, the system must now possess more complexity than its environment *at all scales* rather than merely requiring more complexity than its environment overall.

2.4 A Class of Complexity Profiles

In section 2.3, we have defined the term *complexity profile* and have described general properties that any complexity profile should have. We now describe a specific class of complexity profiles that satisfy these properties. This class of profiles is not the only such class and may not be the best one, but it serves as an instructive example and provides one useful way of characterizing multi-scale complexity.

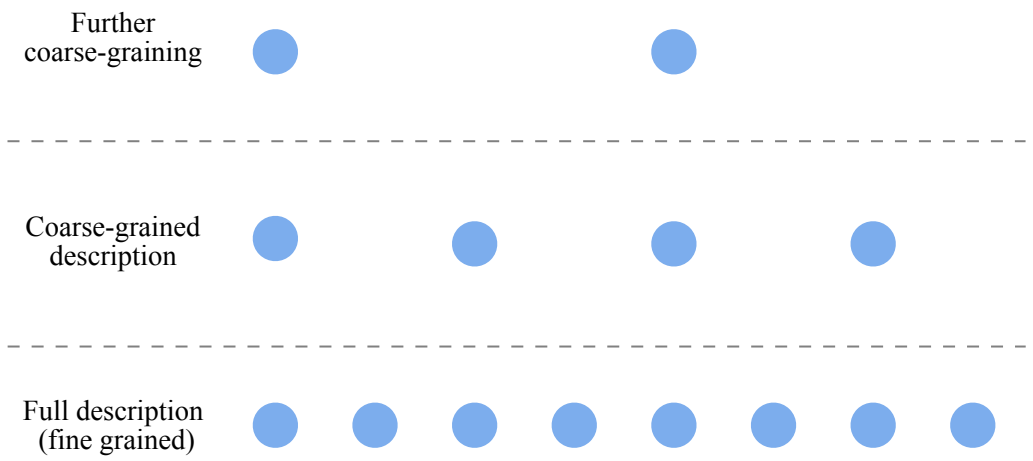


Figure 2-4: Larger-scale/coarse-grained descriptions. Consider a system with eight components. The full description (scale 1) consists of all eight components. A coarse-grained description (scale 2) might consist of every other component, which can serve as an approximation for the system as a whole. A further coarse-grained description (scale 4) might consist of every other component of the scale-2 description.

One way to define a large-scale or coarse-grained description of a system is to allow only a subset of the components of the system to be described.[2] As a first pass, one might divide the system into n equivalent disjoint subsets and then define the information in the description of the system at scale n to simply be the information in one of the subsets (see e.g. fig. 2-4). However, given that the partition into n equivalent subsets may not be possible (either due to heterogeneity in the components or because the system size is not divisible by n), this definition can be generalized by averaging over the the information in each of the n subsets.

Example 5. *Consider a Markov chain* $(x_1, x_2, x_3, ...)$ *(for finite Markov chains of size* N*, simply let* $x_i \equiv 0$ *for* $i > N$*). A set of disjoint, coarse-grained descriptions of the Markov chain at scale* n *could be*

$$\{(x_1, x_{1+n}, x_{1+2n}, x_{1+3n}, ...),\ (x_2, x_{2+n}, x_{2+2n}, x_{2+3n}, ...),\ (x_3, x_{3+n}, x_{3+2n}, x_{3+3n}, ...), ...,\ (x_n, x_{2n}, x_{3n}, ...)\}$$

[2]This is equivalent to the concept of the decimation approach for the position-space renormalization group in physics.

Thus, the information at scale n of the Markov chain could be defined as

$$\frac{1}{n}\sum_{i=1}^{n} H(\{x_j | j \equiv i \mod n\}) \tag{2.2}$$

Note, however, that this sequence of descriptions is not nested, and so cannot be used in its entirety in definitions 8 and 9.

First we must define how to successively partition the system. We only allow for nested sequences of partitions, so that larger-scale descriptions of the system cannot contain information that smaller-scale descriptions lack. The way in which a system is partitioned defines a sequence of descriptions of the system, and thus different partitioning schemes can be thought of as different nested ontologies with which to create successively coarser descriptions of the system. This formulation seeks to allow for a maximally general framework for describing a system at multiple scales, given the constraint that nested, successively larger-scale descriptions of a system correspond to nested subsets of the system that are decreasing in size.

2.4.1 Definition

We now formally define this class of complexity profiles. To do so, we first build up some notation for defining nested sequences of partitions:

Definition 7. *Define $P = \{P_i\}_{i=1}^{\infty}$ to be a* nested partition sequence *of a set X if each P_i is a partition of X, $P_i \leq P_j$ (i.e. P_i is a refinement of P_j) whenever $i > j$, and $P_i < P_j$ (i.e. P_i is a strict refinement of P_j) whenever $|X| \geq i > j$.*

Note that, in order for the strict refinement clause of this definition to be satisfied (i.e. for P_i to have more parts than P_j whenever $|X| \geq i > j$), it must be that P_n contains n parts for $n \leq |X|$ and $P_n = P_{|X|}$ for $n > |X|$, since a partition of X cannot have more than $|X|$ parts.

Example 6. *Let $X = \{x_1, x_2, x_3, x_4\}$. An example of a nested partition sequence of X is*

$$P = (P_1, P_2, P_3, P_4, P_5, ...) = (\{\{x_1, x_2, x_3, x_4\}\}, \quad \{\{x_1, x_3\}, \{x_2, x_4\}\}, \quad \{\{x_1, x_3\}, \{x_2\}, \{x_4\}\},$$
$$\{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}\}, \quad \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}\}, ...)$$

Definition 8. *Given a nested partition sequence P of a system X,*
define $\tilde{S}_X^P(n) \equiv nS_X^P(n) \equiv \sum_{\chi \in P_n} H(\chi)$ for $n \in \mathbb{Z}^+$.

Note that $\tilde{S}_X^P(n)$ is non-decreasing in n and captures the total (potentially overlapping) information of the system parts, while $S_X^P(n)$ functions as a measure of the amount of information necessary to describe the system at scale n, as described above. Information that is n-fold redundant (i.e. is of scale n) can be counted up to n times in $\tilde{S}_X^P(n)$—it is this fact that motivates the following definition of a complexity profile.

Definition 9. *Given a nested partition sequence P of a system X, the complexity profile $C_X^P(n) : \mathbb{Z}^+ \to [0, \infty)$ is defined as $C_X^P(n) = \tilde{S}_X^P(n) - \tilde{S}_X^P(n-1)$, with the convention that $\tilde{S}_X^P(0) = 0$.*

Remark 2. *For $n = 1$, $C_X^P(n) = H(X)$. For $n > |X|$, $C_X^P(n) = 0$. And for $1 < n \leq |X|$, $C_X^P(n) = H(A) + H(B) - H(A, B) = I(A; B)$ where A and B are the two subsets of X that are elements of P_n but not of P_{n-1}. Thus, this complexity profile is very computationally tractable.*

Example 7. *Using the nested partition sequence given in example 6 of $X = \{x_1, ..., x_4\}$, if x_i are unbiased bits, $x_1 = x_2$, and x_1, x_3, x_4 are mutually independent, we have $\tilde{S}_X^P(1) = H(X) = 3$, $\tilde{S}_X^P(2) = H(x_1, x_3) + H(x_2, x_4) = 4$, and $\tilde{S}_X^P(n) = 4$ for $n > 2$. Thus $C_X^P(1) = 3$, $C_X^P(2) = 1$, and $C_X^P(n) = 0$ for $n > 2$.*

Example 8. *Consider a system X of N molecules, the velocities of which are independently drawn from a Maxwell-Boltzmann distribution for which the temperature T is itself a random variable. Consider a nested partitioning scheme in which at each step the largest remaining part (or, in the case of a tie, one of the largest remaining parts) is divided as equally as*

possible in two. The resulting complexity profile will then have $C(1) = H(X)$ *and* $C(n) = H(T)$ *for* $1 < n << N$*, since for* $n << N$*, the size of the parts will be large enough so that* T *can be almost precisely determined from any single part. As* n *approaches* N*, a measurement of any single part will yield more and more uncertainty regarding the value of* T *and so* $C(n)$ *will slowly decay from* $H(T)$ *to* 0*. Such a complexity profile captures the fact that at the smallest scale, there is a lot of information related to the microscopic details of each molecule, but at a wide range of larger intermediate scales, the information present is much smaller and roughly constant, arising only from the common large-scale influence that temperature has across the system.*

2.4.2 The multi-scale law of requisite variety and the sum rule

This complexity profile roughly captures the notion of redundancy and will satisfy the properties described in definitions 5 and 6 (as proved below). It is, of course, dependent on the particular set of partitions used—a reflection of the fact that there is no single way to coarse-grain a system (although some may make more sense than others), and will thus not capture the redundancies present in an absolute sense, as the complexity profile described in refs. [45, 12] does. But that complexity profile, while it does obey the sum rule, does not manifest the multi-scale law of requisite variety. Thus, while it characterizes the information structure present in a system, it does not allow us to compare a system to its environment in a mathematically rigorous way.

Theorem 2.4.1. *Multi-scale law of requisite variety. If a system* X *matches its environment* (Y, f)*, then for all nested partition sequences* P *of* X*,* $C_X^P(n) \geq C_Y^{P^f}(n)$ *at each scale* n*, where* P^f *is the corresponding nested partition sequence of* Y *(see definition 10 below).*

Proof. See section 2.6.1. □

Definition 10. *Given a nested partition sequence* P *of a set* X *and a bijection* $f : Y \to X$*, define* P^f *to be the nested partition sequence of* Y *such that* $\forall n \in \mathbb{Z}^+$*,* $y_1, y_2 \in Y$ *belong to the same part of* P_n^f *iff* $f(y_1), f(y_2) \in X$ *belong to the same part of* P_n*.*

One of the advantages of having the complexity profile depend on the partition is that theorem 2.4.1 holds for all possible nested partitions of the system, assuming its environment is partitioned in the same way. Thus, regardless of how the partitions are used to define scale, the system must have at least as much complexity as its environment at all scales, so long as scale is defined in the same way for the system and its environment.

Furthermore, not only must all possible complexity profiles of the system match the corresponding complexity profile of the environment, but all possible complexity profiles of all possible subsets of the system must match the corresponding complexity profile of the corresponding subset of the environment, as stated in the following corollary to theorem 2.4.1. This is a powerful statement, since it implies that not only must the system have at least as much complexity as its environment at all scales, but also that subdivisions within the system must in some sense be aligned with the natural subdivisions within the environment (see section 1.1.6).

Corollary 2.4.2. *Subdivision matching. Suppose a system X matches its environment (Y, f). Then for any subsets $Y' \subset Y$ and $X' = f(Y') \subset X$, $C^P_{X'}(n) \geq C^{Pf}_{Y'}(n)$ at each scale n for all nested partition sequences P of X'.*

Proof. Since X matches Y, X' matches Y'. Therefore theorem 2.4.1 applies to X' and Y'. □

We now state and prove the sum rule:

Theorem 2.4.3. *Sum rule. For any system X and all nested partition sequences P of X, $\sum_{n=1}^{\infty} C^P_X(n) = \sum_{x \in X} H(x)$*

Proof. $\sum_{n=1}^{\infty} C^P_X(n) = \lim_{n\to\infty} \tilde{S}^P_X(n) - \tilde{S}^P_X(0) = \tilde{S}^P_X(|X|) - 0 = \sum_{x \in X} H(x)$ □

Thus, regardless of which partitioning scheme is used, i.e. regardless of how scale is defined, there is a necessary tradeoff between complexity at larger and smaller scales.

2.4.3 Choosing from among the partitioning schemes

Because of the dependence on the partitioning scheme, definition 9 defines a family of complexity profiles. That there is no single complexity profile for this definition can be thought of as a consequence of their being no single way to coarse-grain a system. In other words, implicit in any particular complexity profile of a system is a scheme for describing that system at multiple scales. While there is no such scheme that is "the correct scheme" in an absolute sense, for any particular purpose (and often for almost any conceivable purpose), some schemes are far better than others.

But before examining this question, we first consider a strong advantage of the multiplicity of complexity profiles: theorem 2.4.1 applies to all of them. Thus, any complexity profile, regardless of the partitioning scheme, can potentially be used to show a multi-scale complexity mismatch between system and environment. This is useful when one has information about the probability distributions of the system and environment separately but not necessarily on the joint probability distribution of system and environment together, such that one cannot directly determine whether the system matches the environment (since quantities such as $H(y|x)$ would be unknown for any given system component x and environmental component y).

Assuming one knows how the system components correspond to the environmental components, one can test for potential incompatibility between the system and environment by considering *any* nested partition sequence of any subset of the system and the corresponding subset of the environment, as per theorem 2.4.1 and corollary 2.4.2. This allows the meaningful comparison of system and environment for a wide variety of definitions of complexity profiles, provided the definitions for the system and environment are consistent. In the likely case that the complexity profiles cannot be precisely calculated, this framework thus supports a wide variety of qualitative complexity profiles that one may wish to construct.

In the case where the correspondence between system and environmental components is either unknown or variable, there are still ways in which to compare the system and the

environment. For instance, if a system X matches its environment Y, then

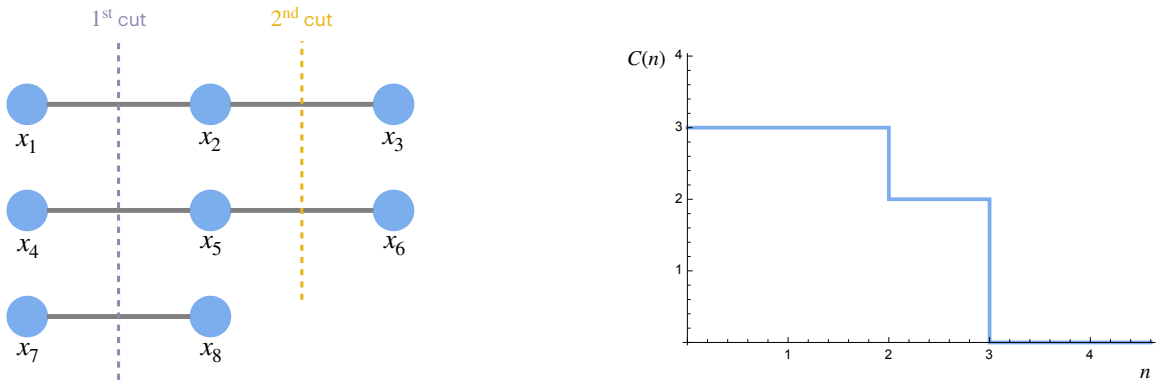
$$\max_P F(C_X^P) \geq \max_P F(C_Y^P) \tag{2.3}$$

for all functions F that map complexity profiles onto $\mathbb{R}$ and are non-decreasing in $C(n)$ for all scales n. Thus, finding even a single function F for which eq. (2.3) does not hold is enough to show that X cannot possibly match Y, regardless of how it may be connected. Other such constructions that are independent of the bijection between X and Y are also possible.

However, although any partitioning scheme can be used to show a mismatch between system and environment, not all partitioning schemes are equally good choices for gaining an understanding of the structure of the system. Each part of a partition represents approximating the system by describing only that subset of its components, and so, if the purpose of the complexity profile is to characterize the structure of the system, the partitions should be chosen accordingly. For instance, for a system $\{x_1, x_2, x_3, x_4\}$ where $x_1 = x_2$ and $x_3 = x_4$, partitioning the system into $\{x_1, x_2\}$ and $\{x_3, x_4\}$ does not make sense if the goal is to create a reasonably faithful two-component description of the four-component system.

As a heuristic, successive cuts in a nested partition sequence should cut through random variables with significant mutual information (i.e. significant redundancy), although, of course, taking a greedy algorithm (i.e.. first maximizing complexity at scale 2, and then choosing the next partition to maximize complexity at scale 3, given the constraint that it has to be nested within the previous partition, and so on) may not always match system structure. Nonetheless, this greedy algorithm does at least provide a consistent way to define complexity profiles across various systems such that complexity is decreasing with scale.[3]

[3]Formally, we can define this complexity profile using the nested partition sequence P that maximizes the complexity profile according to a "dictionary ordering" in which $C_X^{P_1} > C_X^{P_2}$ if there exists an n such that $C_X^{P_1}(n) > C_X^{P_2}(n)$ and for all $m < n$, $C_X^{P_1}(m) = C_X^{P_2}(m)$. Equivalently, this P maximizes $\sum_{n=1}^{\infty} M^{-n} C_X^P(n)$ for any $M > C_X^P(1) = H(X)$. However, just because P maximizes the complexity profile for X according to this (or any other) metric does not guarantee that for an environment (Y, f) of X, $f(P)$ will maximize the complexity profile of Y according to the same metric.



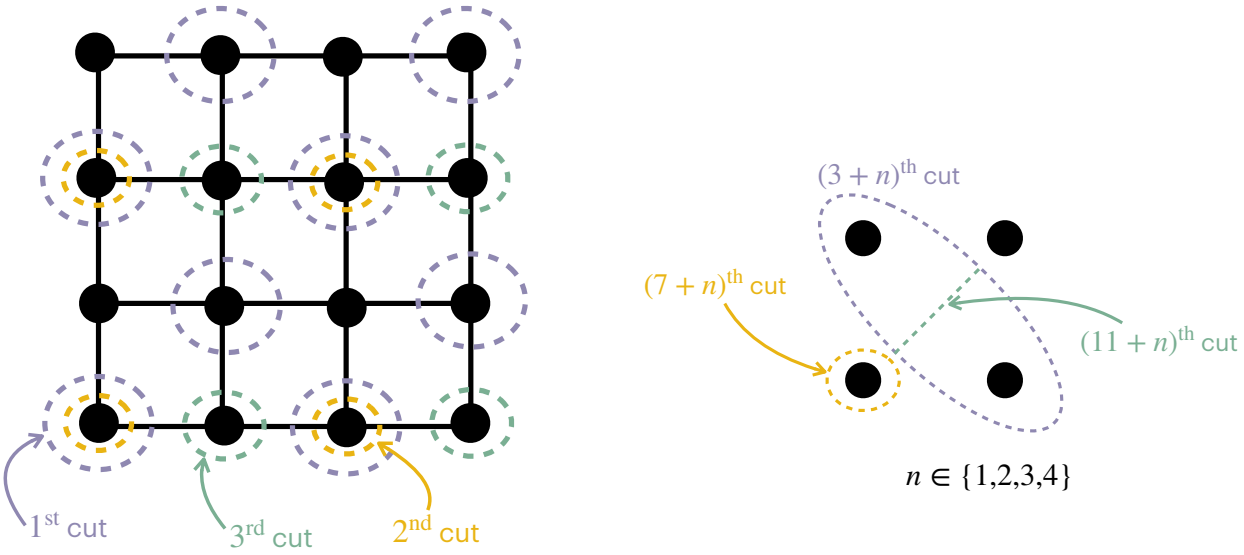
(a) (b)

Figure 2-5: (a) The first and second cuts necessary to create the first three partitions discussed in example 9 are shown, together with (b) the resulting complexity profile (made continuous via eq. (2.1)) if $H(x) = 1$ for each $x \in X$.

Example 9. *Consider a system* $X = \{x_1, x_2, ..., x_8\}$ *such that* $x_1 = x_2 = x_3$, $x_4 = x_5 = x_6$, *and* $x_7 = x_8$, *but otherwise all components are mutually independent (i.e.* x_1, x_4, x_7*) are all mutually independent. Then intuitively we expect that* $C_X(1) = C_X(2) = H(x_1) + H(x_4) + H(x_7)$, $C_X(3) = H(x_1) + H(x_4)$, *and* $C_X(n) = 0$ *for* $n > 3$. *A nested partition sequence that gives us this complexity profile is* $P = (P_1, P_2, P_3, ...)$ *with* $P_1 = \{X\}$, $P_2 = \{\{x_1, x_4, x_7\}, \{x_2, x_3, x_5, x_6, x_8\}\}$, $P_3 = \{\{x_1, x_4, x_7\}, \{x_2, x_5, x_8\}, \{x_3, x_6\}\}$, *and where it does not matter which subsequent partitions are used, since each part of* P_3 *contains mutually independent random variables.*

Example 10. *Consider a two dimensional 4x4 Markov blanket. One way to partition it that respects its structure is given in fig. 2-6.*

Example 11. *Consider a hierarchy consisting of 7 individuals: a leader with two subordinates, each of which have two subordinates themselves, as depicted in fig. 2-7. The behavior of each individual is represented by 3 random variables, each with complexity c. Thus if examined separately from the rest of the system, the complexity of each individual is* 3*c. On the left, everyone completely follows the leader, resulting in a complexity of* 3*c up to scale 7*



(a) (b)

Figure 2-6: Cuts that will create a nested partition sequence for a 4x4 Markov blanket are shown. The first three cuts partition the blanket into four parts with four components each, as shown in (a). Each of the four parts (labeled by $n \in \{1, 2, 3, 4\}$) are then subsequently partitioned as shown in (b) . The resulting complexity profile will of course depend on the nature of the random variables and their correlations.

regardless of the partitioning scheme. On the right, some information is transmitted down the hierarchy but lower levels are also given some autonomy, resulting in more complexity at smaller scales but less at larger scales (but with the same area under the curve, consistent with the sum rule).

2.4.4 Combining subsystems

If a system can be divided into independent subsystems, then the complexity profile of the system as a whole can be written as the sum of complexity profiles of each of the independent subsystems. And if a system can be divided into m subsystems that behave identically, its complexity profile will equal that of any one of the subsystems except with the scale axis stretched by a factor of m. These properties are made precise below.

Theorem 2.4.4. *Additivity of the complexity profiles of superimposed independent systems. Suppose two disjoint systems A and B are independent, i.e. $I(A;B) = 0$, and let $C = A \cup B$. Consider any nested partition sequences P^A of A and P^B of B. Then for all nested partition sequences P^C that restrict to P^A on $A \subset C$ and to P^B on $B \subset C$,*

$$C_C^{P^C}(n) = C_A^{P^A}(n) + C_B^{P^B}(n) \tag{2.4}$$

In other words, the complexity profiles of independent subsystems add.

Proof. This result follows from P_i^C restricting to P_i^A on A and P_i^B on B and the fact that for any subsets $A' \subset A$, $B' \subset B$, $H(A' \cup B') = H(A') + H(B')$. □

In order to formulate the second property, we first build up some notation in the following two definitions:

Definition 11. *For a system X and positive integer m, let $m * X = \cup_{i=1}^{m} X_i$ where the X_i are disjoint systems for which there exist bijections $f_i : X_i \to X$ such that $\forall x \in X_i$, $H(x|f_i(x)) = H(f_i(x)|x) = 0$. In other words, $m * X$ contains m identical copies of X, such that the behavior of any one copy completely determines the behavior of all of the others.*

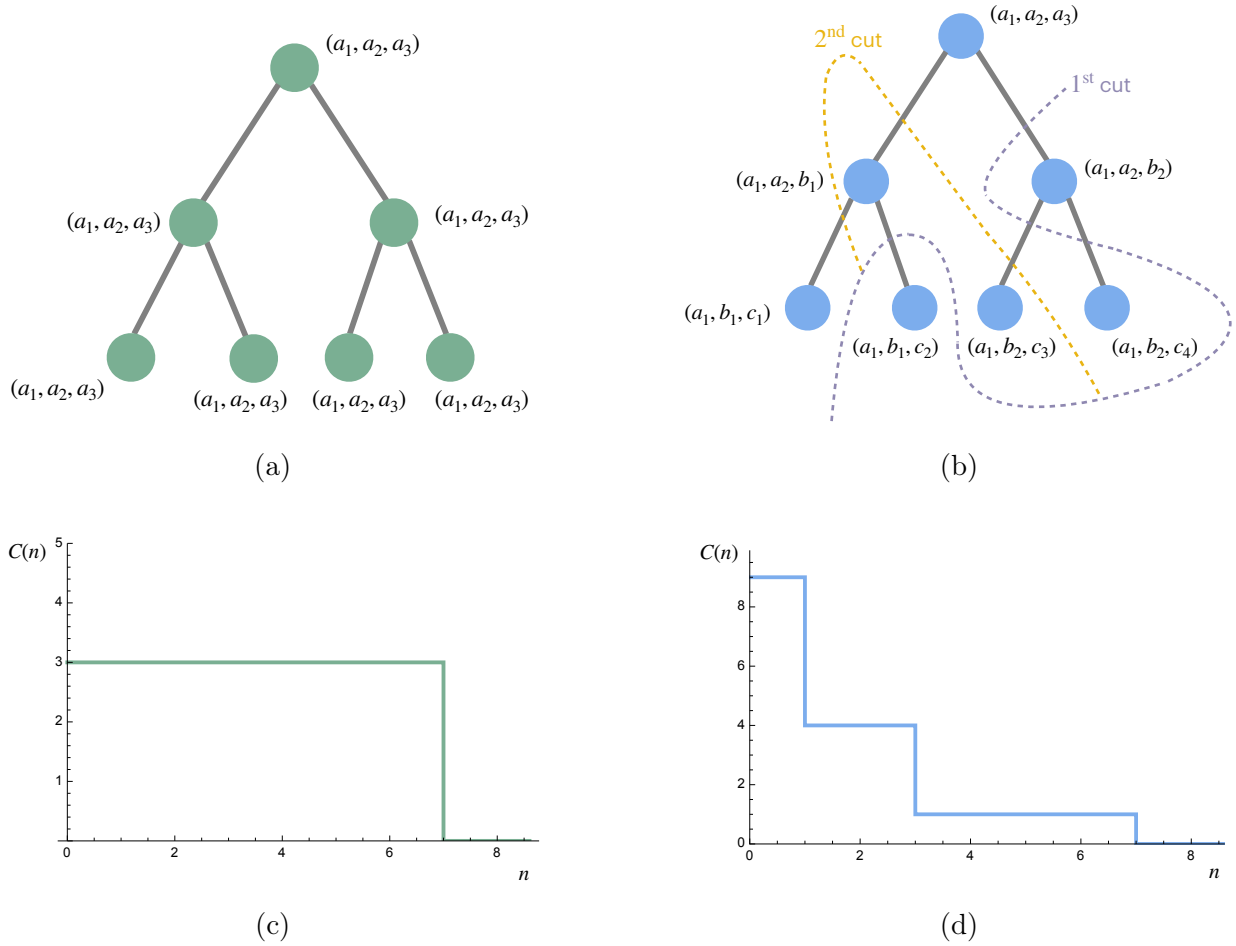


Figure 2-7: The hierarchies discussed in example 11 are shown, together with the resulting complexity profiles (made continuous via eq. (2.1)) for $c = 1$. The complexity profile in (c) can be obtained from any nested partition sequence for the hierarchy in (a). The complexity profile in (d) can be obtained from any nested partition sequence for which the first three partitions are given by displayed cuts in the hierarchy in (b).

Definition 12. *Given a nested partition sequence P of a system X and a positive integer m, define the nested partition sequence $m * P$ of the system $m * X = \cup_{i=1}^{m} X_i$ (with the bijections $f_i : X_i \to X$) as follows. For $n \leq m$, $(m * P)_n \equiv \{\cup_{i=n}^{m} X_i\} \cup \{X_i : i < n\} = \{X_1, X_2, ..., X_{n-1}, \cup_{i=n}^{m} X_i\}$. For $n \geq m$, define $(m * P)_n$ such that it restricts to $P_{n_i}^{f_i}$ (see definition 10) on each $X_i \subset m * X$, where $n_i = \lceil n/m \rceil$ if $i \leq (n \mod m)$ and $n_i = \lfloor n/m \rfloor$ otherwise.*

Theorem 2.4.5. *Scale-additivity of replicated systems. Let P be any nested partition sequence of a system X. Then*

$$C_{m*X}^{m*P}(n) = C_X^P(\lceil n/m \rceil) \tag{2.5}$$

In other words, the effect of including m exact replicas of X is to stretch the scale axis of the complexity profile by a factor of m.

Proof. This result follows from definitions 11 and 12. □

Theorems 2.4.4 and 2.4.5 indicate that for any block-independent system X, i.e. a system in which any two components x_i and x_j are either completely independent (i.e. $I(x_i; x_j) = 0$) or completely dependent (i.e. $I(x_i; x_j) = H(x_i) = H(x_j)$), there exists a nested partition sequence that yields the same complexity profile as that given by the formalism in refs. [12] and [45].[4]

2.5 Conclusion

The motivation behind our analysis here has been to construct a definition of a complexity profile for multi-component systems that obeys both the sum rule and a multi-scale version of the law of requisite variety. In order to do so, we first had to generalize the law of requisite variety to multi-component systems. We then created a formal definition for what constitutes

[4]The formalism in ref. [12]/ref. [45] is stated in ref. [45] to be the only such formalism that is a linear combination of entropies of subsets of the system, that yields its results for block-indpendent systems, and that is symmetric with respect to permutations of the components. The partition formalism in this chapter does not contradict this statement, as the partitions are generally not symmetric with respect to permutations of the system components.

a complexity profile and defined two properties—the multi-scale law of requisite variety and the sum rule—that complexity profiles should ideally satisfy. Finally, we construct a class of examples of complexity profiles and prove that they satisfy these properties. We demonstrate their application to a few simple systems and show how they behave when independent and dependent subsystems are combined.

This formalism is purely descriptive, in that questions of causal influence and mechanism (i.e. what determines the probabilities of the states of each component) are not considered; rather only the possible states of the system and its environment and correlations among these states are considered. By abstracting out notions of causality and mechanism, this approach allows for an understanding of the space of all possible system behaviors and for an identification of systems that are doomed to failure regardless of mechanism. The details by which system and environmental components are mechanistically linked, and the evolution and adaptability of complex systems over time are directions for future work.

More elegant profiles that those presented section 2.4 may exist, and the development of such complexity profiles could be a focus of future work. More broadly, the sum rule could be relaxed, allowing for other definitions of multi-scale complexity. Completely eliminating any tradeoff of complexity among scales would likely lead to under-constrained profiles—certainly, smaller-scale complexity must be reduced in order to create larger-scale structure. But one could imagine other forms that this tradeoff may take other than $\sum_n C_X(n) = \sum_{x \in X} H(x)$. One could also imagine complexity profiles that take advantage of some external structure; for instance, if the system can be embedded into $\mathbb{R}^d$ where d is far lower than the number of system components, Fourier methods could be explored. Even more broadly, other definitions of what it means for a system to match its environment could be considered. But just as the sum rule could be modified but the tradeoff of complexity among scales should nonetheless be manifest, some sort of multi-scale law of requisite variety would need to be present so that the statement that one system has more complexity at a certain scale than another can be meaningful.

With all of that said, the profiles presented here are the first to our knowledge to obey any version of both a sum rule and to manifest any version of a multi-scale law of requisite variety. At the very least, these formalisms provide a formal grounding that can be used to support conceptual claims that are made using complexity profiles. Our hope is that these formalisms spur further development in our understanding of the general properties of multi-component systems, whether these developments are in contrast to or congruent with the formalisms presented here.

2.6 Appendix

2.6.1 Proofs

In order to prove theorem 2.4.1, we first prove the following lemma.

Lemma 2.6.1. *If* $H(B_1|A_1) = H(B_2|A_2) = 0$*, then* $I(A_1; A_2) \geq I(B_1; B_2)$*.*

Proof.

$$
\begin{aligned}
&I(A_1; A_2) = H(A_1) + H(A_2) - H(A_1, A_2) = \\
&H(B_1) + H(A_1|B_1) - H(B_1|A_1) + H(B_2) + H(A_2|B_2) - H(B_2|A_2) \\
&- H(B_1, B_2) - H(A_1, A_2|B_1, B_2) + H(B_1, B_2|A_1, A_2) = \\
&I(B_1; B_2) + H(A_1|B_1) + H(A_2|B_2) - H(A_1, A_2|B_1, B_2) = \\
&I(B_1; B_2) + I(A_1; A_2|B_1; B_2) + I(A_1; B_2|B_1) + I(A_2; B_1|B_2) \geq I(B_1; B_2)
\end{aligned}
$$

□

We now prove **Theorem 2.4.1**: *Multi-scale law of requisite variety. If a system* X *matches its environment* (Y, f)*, then for all nested partition sequences* P *of* X*,* $C_X^P(n) \geq C_Y^{Pf}(n)$ *at each scale* n*.*

Proof. For any collections of random variables $A = \{a_1, .., a_N\}$ and $B = \{b_1, ..., b_N\}$, $\forall i\ H(b_i|a_i) = 0$ implies that $0 \leq H(B|A) \leq \sum_i H(b_i|A) \leq \sum_i H(b_i|a_i) = 0$ and thus $H(B|A) = 0$. Using this fact together with remark 2 and lemma 2.6.1, we get that $C_X^P(n) \geq C_Y^{P^f}(n)$ for $n \in \mathbb{Z}^+$. Note that X and Y being partitioned in the same way guarantees that for a subset $A \subset X$ and the corresponding subset $B \subset Y$, $H(B|A) = 0$. □

2.6.2 Continuum limit

Complexity profiles can also be defined for continuous systems.

Definition 13. *We define a continuous system X of size L as a sequence of discrete systems $\{X_i\}_{i=1}^{\infty}$ with components of size $l_i \equiv L/|X_i|$ such that $X_i \subset X_j$ whenever $i < j$ and $\lim_{i\to\infty} l_i = 0$. Then the complexity profile for the continuous system X is defined as*

$$C_X(s) \equiv \lim_{i\to\infty} \tilde{C}_{X_i}(s) = C_{X_i}(\lceil s/l_i \rceil) \tag{2.6}$$

provided such a limit exists, where $\tilde{C}_{X_i}(s)$ is defined in eq. (2.1).

Remark 3. *Note that any discrete system X with complexity profile $C_X(n)$ and components of size l can be considered as a continuous system of size $|X|l$ per definition 13 by defining the systems $\{X_i\}_{i=1}^{\infty}$ (with components of size l/i) such that $X_i = i * X$ (definition 11) and*

$$\tilde{C}_{X_i}(s) \equiv \tilde{C}_X(s) = C_X(\lceil s/l \rceil) = C_{X_i}(\lceil is/l \rceil) \tag{2.7}$$

(see eq. (2.1)*).*

Example 12. *Suppose that the continuous system X of size L is a random continuous function $f(x)$ for $x \in [0, L]$. Define $X_i = \{f(L/2^i), f(2L/2^i), f(3L/2^i), ..., f(2^iL/2^i)\}$, so that X_i has 2^i components, each of scale $L/2^i$. Then X can be described by the sequence $\{X_i\}_{i=1}^{\infty}$.*

We can extend the class of complexity profiles defined in section 2.4 as follows.

For any nested partition sequence P of a discrete system X with components of size l, the complexity profiles in eq. (2.7) can be realized by letting $C_X(n) = C_X^P(n)$ and $C_{X_i}(n) = C_{X_i}^{i*P}(n)$ (see definition 12), since by theorem 2.4.5,

$$C_{X_i}^{i*P}(\lceil is/l \rceil) = C_X^P(\lceil s/l \rceil) \tag{2.8}$$

To define a partition-based complexity profile using eq. (2.6) for a continuous system X (defined by an infinite sequence of discrete systems $X_1 \subset X_2 \subset X_3 \subset X_4...$, as per definition 13), a nested partition sequence P^i must be chosen for each X_i. Of course, these nested partition sequences must be chosen so that the limit in eq. (2.6) exists; for consistency, it can also be required that on each $X_i \subset X_j$, each partition P_n^j of X_j restricts to P_m^i of X_i for some $m \leq n$.

Chapter 3

Modeling Complex Systems: A Case Study of Compartmental Models in Epidemiology

In this chapter, we use epidemiological compartmental (e.g. SIR, SEIR, etc.) models to illustrate some general modelling principles, and in particular the concept of *effective parameters*.[1]

3.1 Introduction

Compartmental models such as the SIR model have been widely used to study infectious disease outbreaks [171, 280, 55, 262, 60]. These models have informed policy makers of the risks of inaction and have been used to analyze various policy responses. The limitations of the assumptions of compartmental models are well-known [329, 348, 279]; we intend to explore which assumptions are appropriate in which contexts and when and why the models do or do not succeed.

[1]This chapter is taken from the following preprint: Pratyush K. Kollepara, Alexander F. Siegenfeld, Yaneer Bar-Yam. Modeling complex systems: A case study of compartmental models in epidemiology. arXiv:2110.02947 (2021).

No model accurately captures all the details of the system that it represents, but some models are nonetheless accurate because certain large-scale behaviors of systems do not depend on all these details (see section 1.2.4). (For example, modeling material phase transitions generally does not require including the quantum mechanical details of individual atoms.) The key to good modeling is understanding which details matter and which do not. Paradoxically, failing to recognize that a model can be accurate in spite of certain unrealistic assumptions can lead to models in which all assumptions are excused: the impossibility of getting all the details right may discourage a careful analysis of which assumptions are appropriate in which contexts.

During a pandemic, it is crucial that models complement decision-making. In an attempt to obtain better predictions, it may be tempting to include more details and fine-tune the model assumptions. However, focusing on irrelevant assumptions and details while losing sight of the large scale behavior is counterproductive [300]. Which details are relevant depends on the question at hand; the inclusion or exclusion of details in a model must be justified depending on the modeling objectives. Compartmental models tend to include some details (e.g. disease stages) while not including others (e.g. stochasticity and heterogeneity) that, in many cases, have a far larger effect on forecasting the epidemic trajectory, estimating the final epidemic size, and analyzing the impact of interventions.

In this chapter, we examine some common assumptions of compartmental models—such as the distribution of generation intervals, homogeneity in population characteristics and connectivity, and the use of continuous variables—in order to determine their relevance for various model outcomes. Our purpose is not to argue for specific alternatives to compartmental models or for specific modifications but rather to illustrate how the assumptions of these models affect their results.

3.2 The SIR model

Here we introduce the SIR model. The model divides the population into three compartments—the fractions of individuals who are susceptible (s), infected (i), or recovered (r). A set of

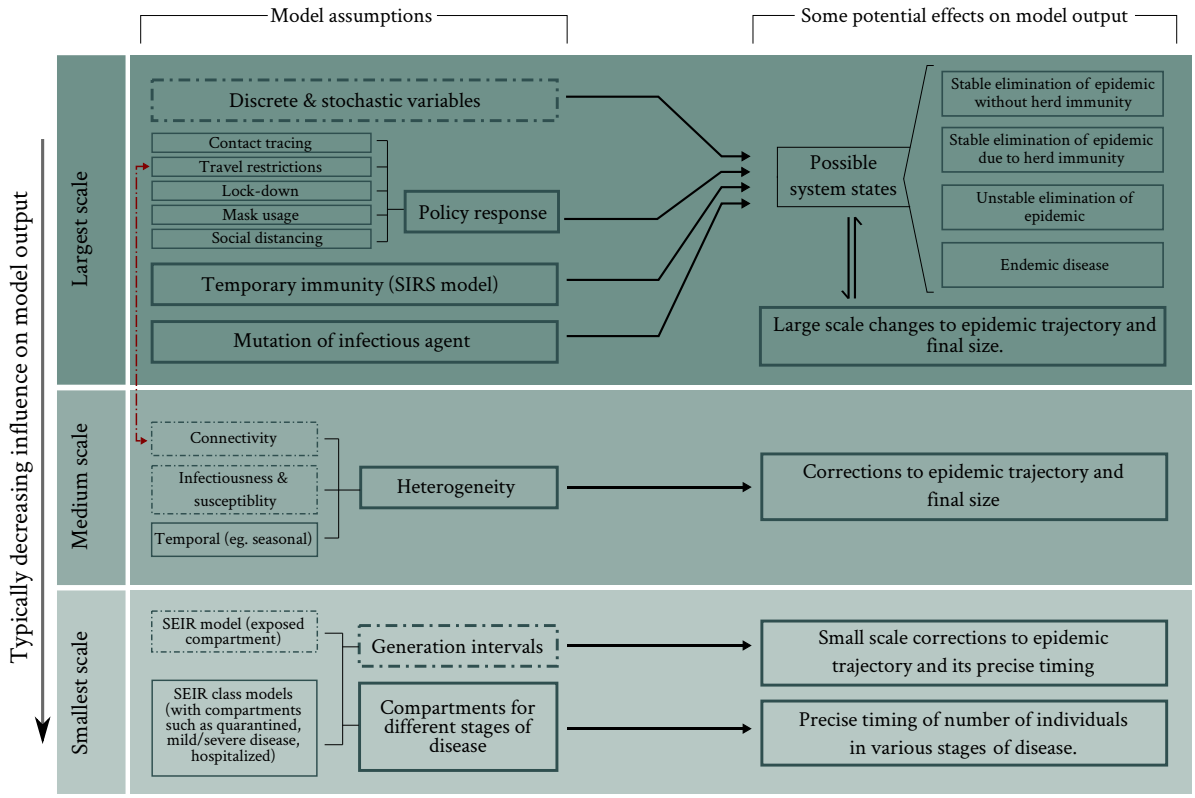


Figure 3-1: Schematic representation of the impact of various modeling choices/assumptions. The left column lists various details that can be incorporated into a compartmental model (boxes with dashed borders indicate modeling choices that are analyzed in section 3.3), and the right column lists typical potential impacts on the model output. The three panels classify the system details by 'scale', with the largest scale details typically having the most impact on model output, and the smallest scale details typically having the least impact, although the impact of any given assumption ultimately depends on precisely for what purpose the model is being used. For instance, an SIRS model may not be needed if only the initial growth of the epidemic is being modeled. Furthermore, various assumptions can compound non-linearly to affect the model output. For instance, policy interventions such as travel restrictions, which both rely on and affect heterogeneity in geographical connectivity, can play a decisive role in determining whether or not a stable elimination is achieved [297]. Of course, the actual effect of any assumption depends on its precise mathematical implementation, as well as the presence or absence of other assumptions within the model, and so this figure should be considered as a rough schematic rather than as a definitive guide.

three differential equations governs the dynamics:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta si \tag{3.1}$$

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta si - \gamma i \tag{3.2}$$

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \gamma i \tag{3.3}$$

The parameter $\beta > 0$ is the rate at which an infected individual transmits the disease to a susceptible individual. The infected individuals become no longer infectious (recovered or removed) at a rate $\gamma > 0$. Assumptions of the SIR model include homogeneity in the infectiousness, susceptibility, and connectivity of the population, exponentially distributed recovery times and generation intervals, that discrete and stochastic dynamics can be approximated with continuous and deterministic variables, and that there are no changes over time in the behaviors of either the population or the infectious agent.

By recasting the equations of the model in terms of the basic reproduction number $R_0 = \beta/\gamma$,

$$\frac{\mathrm{d}s}{\mathrm{d}(\gamma t)} = -R_0 si \tag{3.4}$$

$$\frac{\mathrm{d}i}{\mathrm{d}(\gamma t)} = (R_0 s - 1)i \tag{3.5}$$

$$\frac{\mathrm{d}r}{\mathrm{d}(\gamma t)} = i \tag{3.6}$$

it can be seen that the evolution of the system state (i.e. the fraction of people in each of the three compartments) depends only on R_0 and that γ sets the time scale for this evolution (i.e. a change in γ would correspond simply to a stretching or compression of the time axis). Indeed, it can be proven that the final size of an epidemic depends only on the network of probabilities of individuals infecting each other and not at all on how quickly individuals recover or any other time-scales associated with the progression of the disease within individuals [248, 17, 231, 35, 54] (see Appendix section 3.5.1 for more details).

This overall time-scale of the epidemic (set by γ in the above formulation) is an important parameter; for instance, together with R_0, it tells us how quickly case counts will grow. The SIR model describes this overall time-scale without the need for any additional compartments. Studies with additional compartments that have focused on the details of infectious periods, latent periods, and other disease stages can provide more information as to the precise timing (as opposed to simply the overall fraction) of the number of individuals in particular disease stages (e.g. exposed, infectious, hospitalized, etc.) if sufficient data is available to fit the additional parameters. This approach is useful for understanding, for instance, lags between infections and hospitalizations. However, such details will often have much smaller effects than regularly used assumptions that impact the overall epidemic trajectory, such as homogeneity, mean-field connectivity, and continuous variables (section 3.3).

3.3 Analysis of key assumptions

We now examine some key assumptions of compartmental models. In section 3.3.1, we show that assumptions concerning the distribution of generation intervals (i.e. assumptions about diseases stages, recovery rates, etc.) do not significantly affect the overall epidemic trajectory. In section 3.3.2 we show how the effects of heterogeneity in susceptibility and connectivity cannot be captured by an average or effective spreading rate β, in contrast to how the distribution of generation intervals can be described by the effective parameter γ. In section 3.3.3, we discuss the implications of using continuous and deterministic variables to describe dynamics that are in reality stochastic and discrete.

3.3.1 SEIR models, generation intervals and effective parameters

There are some cases in which simplifying assumptions are not critical. For instance, by assuming a constant recovery rate γ, the SIR model makes the assumption that generation intervals follow an exponential distribution. However, the growth rate and reproduction

number can nonetheless be accurately captured despite the actual generation intervals not being exponentially distributed, so long as γ is treated as an effective parameter. Given an observed reproduction number R_0 and initial exponential growth rate λ_0, one may always find a β and γ such that $R_0 = \beta/\gamma$ and $\lambda_0 = \beta - \gamma$. We note in this case that $1/\gamma$ need not be the mean of the distribution of generation intervals $g(t)$. Instead, the relationship between γ, λ_0, and R_0 is given by

$$\frac{1}{R_0} = \int_0^\infty g(t)e^{-\lambda_0 t}dt \tag{3.7}$$

which implies that the γ necessary to match R_0 and λ_0 will be the inverse mean of $g(t)$ if and only if $g(t) = \gamma e^{-\gamma t}$ [343] (see Appendix section 3.5.2 for details).

Models with additional compartments such as SEIR models are often considered to be more accurate than the SIR model since they include a more realistic generation interval distribution (see Appendix section 3.5.3). However, the precise generation interval distribution does not affect epidemic characteristics such as the final size, the initial exponential growth rate, and R_0. As described above, these characteristics can be captured by the SIR model by treating the recovery rate γ as an effective parameter (see Figure 3-2). More generally, for the purposes of modelling the overall epidemic trajectory, introducing any number of disease stages into the SIR model only amounts to changing the effective distribution of generation intervals, which changes only the precise timing of the epidemic curve (see Appendix section 3.5.1). The SIR model is elegant in that its parameter set (the dimensionless R_0 plus a time scale γ) is minimal; given the larger sources of uncertainty related to other assumptions, additional parameters in SEIR models are not justified if they serve only to refine the generation interval distribution. (The use of SEIR models over the SIR model may be justified in other circumstances.)

SIR, SEIR and other compartmental models are frequently used for estimating the basic reproduction number [60], but very often, the transition rates are not treated as effective parameters and are estimated using the inverse of mean generation, serial, latent or infectious intervals [90, 180, 206, 276, 324, 357, 364], which, as described above, is appropriate only

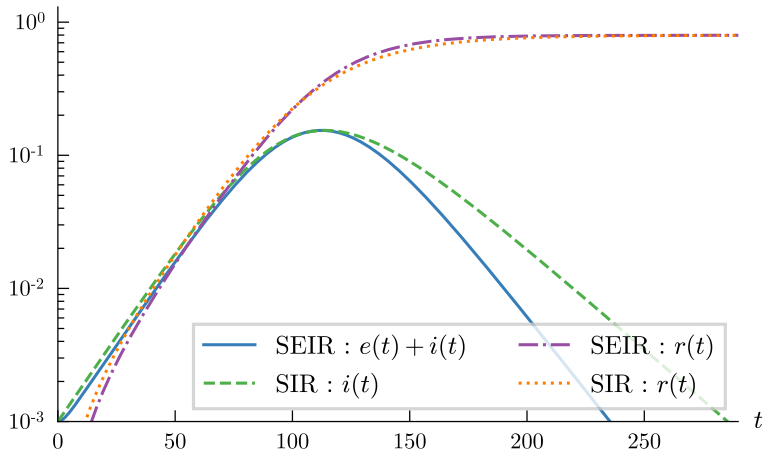


Figure 3-2: An SEIR model can be replaced with an SIR model with a nearly identical trajectory. Important characteristics such as the growth rate, reproduction number, herd immunity threshold and epidemic size will be exactly the same in both models. The following parameters were used to generate this figure. SEIR: $\beta = 0.25$, $\sigma = 0.167$, $\alpha = 0.125$. SIR: $\bar{\beta} = 0.119$, $\gamma = 0.059$

for exponentially distributed intervals. This practice is also prevalent in other epidemic modeling literature [207, 342, 111, 92, 283, 287, 30, 31, 32, 89, 122, 137].

3.3.2 Population heterogeneity

Human populations are heterogeneous in many ways: social networks of individuals exhibit community structure [151, 18, 167], infectiousness and susceptibility can vary across the population depending upon age/health conditions/behavior, different regions may have different mitigation responses to an epidemic, etc. Therefore, in this section we will discuss the widely used assumption of homogeneous and well-mixed populations [207, 342, 111, 92, 137]. The homogeneity assumption has been challenged using various types of heterogeneous models [72, 248, 157, 145, 172, 121], and these studies point towards a crucial result: heterogeneous models can yield very different outcomes from homogeneous models.

To summarize this impact of heterogeneous infectiousness, susceptibility, and connectivity, we use a simple class of models in which the population is partitioned into multiple groups [72, 248, 157, 145, 172, 121]. We briefly describe these models below (methods can be

found in Appendix section 3.5.4). The purpose of these modifications is not to create a more accurate model but to show that heterogeneity and connectivity are crucial assumptions that can have a substantial effect on both the epidemic trajectory and its final size. We do not claim that any particular set of assumptions regarding heterogeneity or connectivity will accurately predict an epidemic trajectory but rather include such assumptions to show that the space of possible outcomes is far larger than homogeneous models would imply.

Heterogeneity in infectiousness and susceptibility

The SIR model assumes that individuals spread infections in a homogeneous manner. We consider three ways in which spreading can be heterogeneous. Figure 3-3 summarizes the results of these modifications to the model and shows that the final size can be very different despite identical initial exponential growth rates and basic reproduction numbers.

First, groups can be equally susceptible (without loss of generality, $\eta = 1$) but differ in infectiousness β. In this case, the SIR model can effectively coarse-grain this heterogeneity by selecting the effective spreading rate $\beta_{\text{SIR}} = \langle\beta\rangle$, where the angled brackets indicate an average over the whole population.

Second, the groups can have the same infectiousness $\bar{\beta}$ but differing susceptibilities η. By selecting the effective spreading rate $\beta_{\text{SIR}} = \bar{\beta}\langle\eta\rangle$, the initial growth rate and basic reproduction number can be reproduced with an SIR model. However, the SIR model will offer a substantially different prediction for later parts of the epidemic trajectory and the final epidemic size.

In the third case, both infectiousness and susceptibility vary across groups. We consider a special case of this scenario by assuming that infectiousness and susceptibility are proportional, i.e. those who are more likely to spread the disease are also more likely to contract it. For instance, a person who wears a mask more often or who socializes less will be both less likely to spread and less likely to contract the disease. Assuming both susceptibility and infectiousness are proportional to a contact parameter b, the homogeneous SIR model can reproduce the initial growth rate and basic reproduction number by selecting the effective

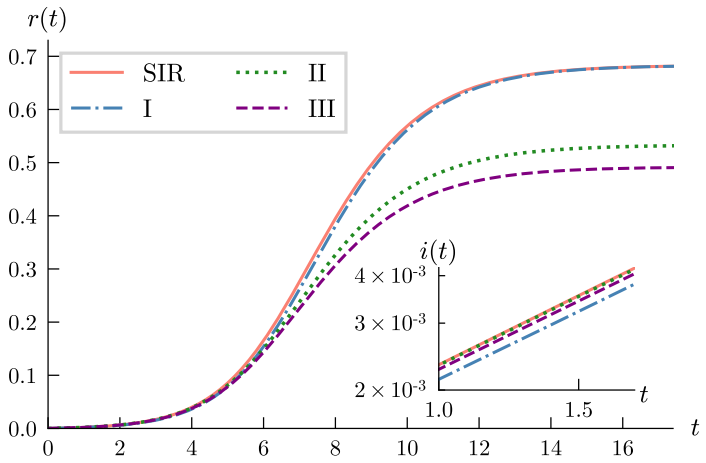


Figure 3-3: Effect of different types of heterogeneity on the trajectory of an epidemic. (I) Heterogeneity in infectiousness. (II) Heterogeneity in susceptibility. (III) Heterogeneity in both infectiousness and susceptibility. All four cases have the same value of γ and the same initial growth rate (as shown in the inset). Despite similar initial behavior, (II) and (III) exhibit different epidemic sizes (the saturated value of $r(t)$). The equations in Appendix section 3.5.4 were used to generate the trajectories.

spreading rate $\beta_{\text{SIR}} = \langle b^2 \rangle$. (Note that here, the effective spreading rate $\langle b^2 \rangle$ differs from the average spreading rate $\langle b \rangle^2$, due to the more infectious individuals being more likely to be infected.) However, as in the previous case, the homogeneous SIR model can grossly misestimate later parts of the trajectory and the final epidemic size.

These results show that unless all individuals are equally likely to be infected, the large-scale effects of heterogeneity on the epidemic trajectory beyond the initial exponential growth cannot be captured by a homogeneous model. In other words, the heterogeneity cannot be coarse-grained into a single effective parameter.

Heterogeneous connectivity

The SIR model assumes mean-field connectivity (i.e. every individual is equally likely to interact with every other individual). Here we consider a population for which the connectivity within and between groups can be controlled through a clustering parameter between zero and one. A clustering parameter of zero means that the groups are perfectly well-mixed while a clustering parameter of one means that there is no inter-group interactions

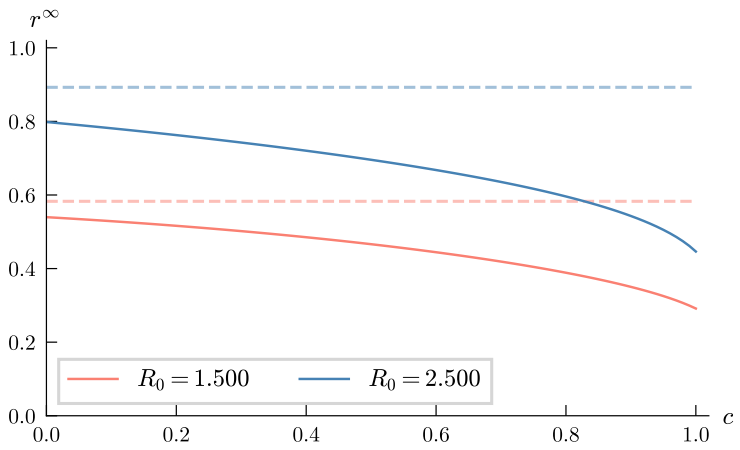


Figure 3-4: Effect of heterogeneous connectivity, infectiousness, and susceptibility on the final epidemic size (r^∞). For a given value of R_0, varying the clustering parameter c from 0 to 1 (and adjusting the contact parameter b_1 so as to maintain the same value of R_0) in a population containing two groups can lead to epidemics of different sizes. The dashed curves of corresponding colors show the epidemic size for the same value of R_0 in the homogeneous SIR model. Parameter values are $b_2 = 0.9, n_1 = n_2 = 0.5, \gamma = 1.0$ (see equation 3.42) .

(mathematical methods can be found in Appendix section 3.5.4). Figure 3-4 shows how connectivity assumptions can affect epidemic size. More important, however, is the space of policy responses that is opened up by the fact that connectivity is not mean-field (i.e. that populations are not well-mixed). The geographic clustering of cases, which can be increased with travel restrictions, can be especially helpful in containing a pandemic using only local, targeted measures [297].

Another large-scale effect of heterogeneity is that the trajectory of the number of infections can have multiple peaks, an impossible occurrence under homogeneous compartmental models (see Figure 3-5). Of course, the shape of an epidemic trajectory will also be affected by policy interventions, behavioral changes in the population, evolution of the infectious agent, and seasonal effects, as well as nonlinear interactions among and between these factors and various heterogeneities. In Figure 3-6 (see Appendix), we discuss an observed plateau in the epidemic time-series from India that can be partially explained on the basis of heterogeneity.

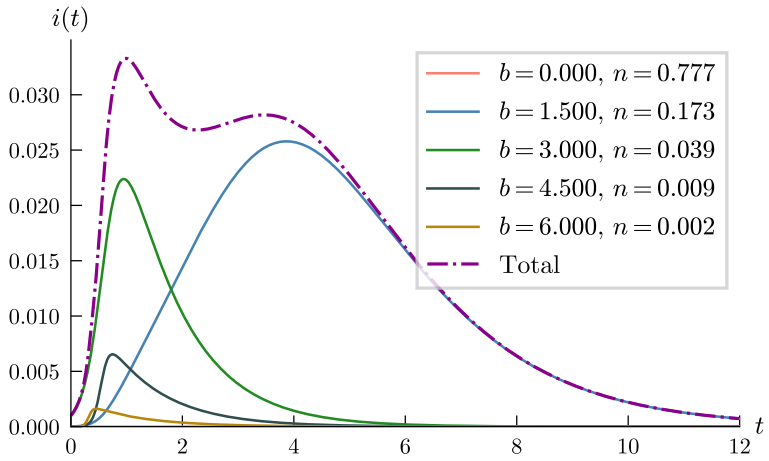


Figure 3-5: Heterogeneous connectivity, susceptibility, and infectiousness can substantially change the trajectory of the epidemic. Since the groups are well separated, each group exhibits a unique growth rate. If a homogeneous compartmental model was used to forecast the trajectory at $t \sim 1.5$, we would be led to believe that the epidemic was about to end. Parameters: $c = 0.75$, $\gamma = 1$, number of groups $= 5$, contact parameters b are approximately exponentially distributed with mean 1. Seed infection is in the group with $b = 3$ (see Appendix section 3.5.4).

3.3.3 Continuous variables and elimination of outbreaks

In compartmental models, which use continuous variables, the number of infections can exponentially decay but will never reach zero. Such models may mischaracterize the effects of temporary, strong interventions by predicting an inevitable "second wave" [342, 111, 137]. Stochastic compartmental models [13] are better suited for analyzing interventions since they use discrete variables and present elimination as a possible scenario (see Appendix section 3.5.5 for details).

Stochastic models also show that not all outbreaks grow to become an epidemic [15], an observation which can aid in identifying policies that achieve containment once cases have been brought to a sufficiently low number. The effect of stochasticity can be particularly pronounced if super-spreader events play a substantial role in the overall spread of the disease.

Since stochastic disease transmission events take place through the contact networks of individuals, connectivity patterns can affect the dynamics of elimination. Suspending

long-range connections (through lock-downs and travel restrictions) can lead to localized epidemics that are largely independent of each other. These local epidemics will be smaller and thus able to be more quickly eliminated. Thus, heterogeneous connectivity can interact with stochastic effects to make elimination a more accessible prospect than homogeneous models would imply.

3.4 Discussion

What differentiates a good model from a bad model is not its level of detail but rather the relationship between the details included in the model and the most important behaviors of the system. Which details are important can depend not only on the system but also on the modelling objectives. For instance, the precise distribution of generation intervals is crucial if attempting to calculate the reproduction number from the exponential growth/decline rate of an epidemic, but it can generally be coarse-grained to a single time-scale in the context of predicting overall epidemic trajectories.

SEIR models differ from the SIR model in that they use a more detailed and realistic generation interval distribution. The corrections to the epidemic trajectory from the distribution of generation intervals will generally be small compared to other sources of error. Both SIR and SEIR models ignore potentially important factors such as heterogeneity, stochasticity, and behavior change/policy interventions. If such factors are to be ignored, however, the SIR model has the advantage of not including any unnecessary (and therefore potentially misleading) details; its output depends only on the unit-less R_0, together with a time-scale set by the effective recovery rate γ.

SEIR and related models can provide a breakdown of types of infections and can help with, for example, the management of health care resources (although often, all that matters is the probability of an infection being of a certain type rather than the precise dynamics between types). However, given the often far larger effects of heterogeneity and stochasticity (not to mention behavioral change and policy response), including the details of disease progression while ignoring these other assumptions may provide a false sense of confidence in

the accuracy of the model. More importantly, a misunderstanding of the relative importance of assumptions in any given model may narrow the set of interventions considered.

The idea that some details and assumptions are more important than others is frequently used in mathematics and physics. In mathematics, functions are often approximated locally using a Taylor expansion, with each higher order term providing additional details. A higher order term (finer-grain correction) is used only when all lower-order terms (coarser-grain corrections) have been included. To do otherwise—or to include some corrections at a given scale while ignoring others—is fundamentally unsound and can lead to nonsensical results. When modeling a real-world system, the various details do not necessarily fit cleanly into a Taylor expansion, but the general conceptual principle still holds: details of lower relative importance should be considered only after all of the larger-scale effects have already been taken into account.

Agent-based or network models can transcend some of the limitations of compartmental models. Like any model, however, they may suffer from the flaw of arbitrarily focusing on some details while leaving out others, thereby mischaracterizing the space of large-scale behaviors of the epidemic. As agent-based and network models are generally more detailed, especially careful attention must be paid to this point.

We have not examined many details such as contraction of generation intervals [195], policy responses to an outbreak, temporal and geographic heterogeneity, time varying immunity, seasonal effects, and the emergence of disease variants, among others, which may influence the large-scale behaviors of an epidemic in dramatic ways. The unpredictability inherent to epidemics underscores the need for a precautionary principle for acting under uncertainty [95]. A careful examination of model assumptions is needed, not to evaluate the accuracy of the assumptions themselves—they will always be inaccurate—but to see how they do or do not affect the link between our actions and the space of possible outcomes.

3.5 Appendix

3.5.1 Epidemic size

The epidemic size is not affected by the generation interval distribution. This can be seen by considering an epidemic process on a graph of N nodes, where an infected individual has a chance p of infecting any other susceptible, and that these infection events are independent of each other. The infected contacts made in this process constitute an Erdos-Renyi graph [54, 35]. Thus, size of the epidemic is the size of the giant component of the Erdos-Renyi graph, the expectation value of which is given by $z = 1 - \exp(-pNz)$, which is identical to the epidemic size of the SIR model if $R_0 = pN$. Under the assumptions of homogeneity, the epidemic size depends only on R_0, and not on the generation interval distribution, spreading or recovery rates.

More generally, consider a set of nodes on which each transmission event leads to a directed edge. The transmission events need not be independent (to account for super spreading behavior or heterogeneity). The final size of the epidemic is the number of nodes that are connected directly or indirectly to the seed infection node. So, the final size depends only on the probabilities of transmission (existence of a directed edge between a pair of nodes), and does not depend on the temporal properties of the process such as generation intervals or the number or types of stages in a compartmental model.

3.5.2 Generation intervals

The SIR model, despite its unrealistic assumptions about the generation interval distribution, can correctly capture the initial exponential growth rate, the basic reproduction number R_0, and the epidemic size (subject to the assumption of homogeneity).

A general result is that if the generation intervals of an epidemic are given by a random variable T, the relationship between the effective reproduction number R and growth rate is

given by

$$R = 1/M_T(-\lambda) \tag{3.8}$$

where $M_T(\mu) = \mathbb{E}[e^{\mu T}]$ is the moment generating function of the distribution of T and λ is the exponent of growth or decline [343]. This relationship is valid for any distribution of generation intervals and applies whenever the population size and number of infections is large enough that stochastic effects can be ignored. (There is also the implicit assumption that R and the generation interval distribution are roughly constant over the time interval during which exponential growth/decline is observed.)

Within an SIR model, the generation intervals are exponentially distributed with mean $1/\gamma$, so equation (3.8) yields

$$R_0 = 1 + \lambda_0/\gamma \tag{3.9}$$

where λ_0 is the initial exponential growth rate. If an SIR model is to accurately describe R_0 and λ_0 for an observed epidemic, then γ is determined by Equation (3.9). But since actual generation intervals are not exponentially distributed, the inverse recovery rate $1/\gamma$ can not be estimated as the mean of the observed generation intervals. Instead, γ (and β) serve as effective parameters that coarse-grain the actual distribution of generation intervals T in such a way that the SIR model yields the correct initial growth rate λ_0 and basic reproduction number R_0:

$$\gamma = \frac{\lambda_0}{R_0 - 1} = \frac{\lambda_0}{1/M_T(-\lambda_0) - 1} \neq \frac{1}{\mathbb{E}[T]} \tag{3.10}$$

$$\beta = \gamma R_0 = \frac{\lambda_0}{1 - M_T(-\lambda_0)} \tag{3.11}$$

Nonetheless, much of the modeling literature (e.g. [90, 180, 206, 276, 324, 357, 364, 207, 342, 111, 92, 283, 287, 30, 31, 32, 89, 122, 137]) uses $\gamma = \frac{1}{\mathbb{E}[T]}$.

3.5.3 The SEIR model

The SEIR model is an enhanced SIR model where a new compartment of exposed individuals (who have been infected but are not infectious) is introduced. Exposed individuals transition to the infectious compartment at a rate σ. The model is described these differential equations:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta si \tag{3.12}$$
$$\frac{\mathrm{d}e}{\mathrm{d}t} = \beta si - \sigma e \tag{3.13}$$
$$\frac{\mathrm{d}i}{\mathrm{d}t} = \sigma e - \alpha i \tag{3.14}$$
$$\frac{\mathrm{d}r}{\mathrm{d}t} = \alpha i \tag{3.15}$$

Convolving the two exponential distributions corresponding to the transitions from exposed to infectious and infectious to recovered [343] gives the following distribution of generation intervals

$$g(t) = \frac{\sigma\alpha}{\sigma - \alpha}(e^{-\alpha t} + e^{-\sigma t}) \tag{3.16}$$

which, when combined with equation (3.8), yields

$$R_0 = (1 + \lambda_0/\sigma)(1 + \lambda_0/\alpha) \tag{3.17}$$

where λ_0 is the initial exponential growth rate and R_0 is the basic reproduction number. Combining equations (3.17) and (3.9) allows us to find an effective γ of the SIR model in terms of the effective parameters of the SEIR model such that the basic reproduction number and the initial growth rate of both the models are equal:

$$R_0 = (1 + \lambda_0/\sigma)(1 + \lambda_0/\alpha) = 1 + \lambda_0/\gamma \tag{3.18}$$

Thus, any SEIR model can be replaced with an SIR model with the same initial growth rate and R_0 (and thus final size). (Note that for both models, β is also an effective parameter:

for the SEIR model, $\beta = R_0\alpha$, while for the SIR model, $\beta = R_0\gamma$) The two models will differ only in terms of the precise timing the epidemic curve later on in its trajectory, but such differences will be swamped by other sources of error such as heterogeneity and stochasticity.

3.5.4 Heterogeneity

Heterogeneity in infectiousness, susceptibility and connectivity that were explored in section 3.3.2 use a common framework in which a population of size N is divided into multiple groups. For a group k: size of the group is N_k and the density is $n_k = N_k/N$. The number of susceptible, infected and recovered individuals is S_k, I_k, R_k respectively and their corresponding densities are $s_k = S_k/N, i_k = I_k/N, r_k = R_k/N$, such that $s_k + i_k + r_k = n_k$. For two groups k and l, the inter group interactions (infectiousness, susceptibility and connectivity) are captured by B_{kl}. The modified SIR equations can be written as:

$$\frac{\mathrm{d}i_k}{\mathrm{d}t} = s_k \sum_l B_{kl} i_l - \gamma i_k \tag{3.19}$$

$$\frac{\mathrm{d}s_k}{\mathrm{d}t} = -s_k \sum_l B_{kl} i_l \tag{3.20}$$

$$\frac{\mathrm{d}r_k}{\mathrm{d}t} = \gamma i_k \tag{3.21}$$

It can be seen that a system state with no infections ($i_k^* = 0$ for all k) is a fixed point of these equations. At the beginning of the epidemic, the number of infections in each group will be approximately zero ($i_k \approx 0$) and most of the population will be susceptible ($s_k \approx n_k$). Thus equation (3.19) can be linearized about the fixed point $i_k^* = 0$ to give

$$\frac{\mathrm{d}i_k}{\mathrm{d}t} = n_k \sum_l B_{kl} i_l - \gamma i_k \tag{3.22}$$

This is a set of linear differential equations whose initial time behaviour can be understood from the eigenvalues of the matrix M where $M_{kl} = n_k B_{kl} - \gamma\delta_{kl}$. More precisely, the top eigenvalue of M gives the initial exponential growth rate of number of infections. As the

epidemic progresses, the fraction of infected population is no longer close to zero and the epidemic deviates from exponential growth.

The basic reproduction number is the top eigenvalue of the next generation matrix [118, 338]. The next generation matrix G is given by

$$G_{kl} = \frac{n_k B_{kl}}{\gamma} \tag{3.23}$$

The next generation matrix can also be obtained from the matrix M using $G = 1 + \frac{1}{\gamma}M$. Thus the basic reproduction number can be computed using equation (3.9).

The size of the epidemic $r_k(t \to \infty)$ can be estimated by using equations (3.20) and (3.21), integrating from $t = 0$ to ∞, with the condition that $i_k(t \to \pm\infty) = 0$.

$$r_k^\infty = n_k \left[1 - \exp\left\{ -\frac{1}{\gamma} \sum_l B_{kl} r_l^\infty \right\} \right] \tag{3.24}$$

$$r^\infty = \sum_k r_k^\infty \tag{3.25}$$

Heterogeneity in individual characteristics

The following equations were used for the three cases described in section 3.3.2 of the main text and give expressions for the initial exponential growth rate λ_0, the basic reproduction number R_0 and final epidemic size.

Case I: Heterogeneous infectiousness and homogeneous susceptibility ($B_{kl} = \beta_l$).

$$\lambda_0 = \sum_l n_l \beta_l - \gamma \tag{3.26}$$

$$= \langle \beta \rangle - \gamma \tag{3.27}$$

$$R_0 = 1 + \lambda_0/\gamma = \langle \beta \rangle / \gamma \tag{3.28}$$

One can show that eq. (3.24) is satisfied if $r_l^\infty = n_l r^\infty$, and thus

$$\frac{r_k^\infty}{n_k} = 1 - \exp\left\{-\frac{\langle\beta\rangle r^\infty}{\gamma}\right\} = r^\infty \tag{3.29}$$

$$r^\infty = r_{\mathrm{SIR}}^\infty \tag{3.30}$$

Thus, an SIR model with an effective spreading rate $\beta_{\mathrm{SIR}} = \langle\beta\rangle$ not only reproduces the initial growth rate and basic reproduction number but also the final size (and, as it can be shown with similar logic, the entire epidemic trajectory).

Case II: Heterogeneous susceptibility and homogeneous infectiousness ($B_{kl} = \bar{\beta}\eta_k$)

$$\lambda_0 = \bar{\beta}\sum_k n_k\eta_k - \gamma \tag{3.31}$$

$$= \bar{\beta}\langle\eta\rangle - \gamma \tag{3.32}$$

$$R_0 = 1 + \lambda_0/\gamma = \bar{\beta}\langle\eta\rangle/\gamma \tag{3.33}$$

$$\frac{r_k^\infty}{n_k} = 1 - \exp\left\{-\frac{\bar{\beta}\eta_k}{\gamma}\sum_l r_l^\infty\right\} \tag{3.34}$$

By selecting the effective spreading rate $\beta_{\mathrm{SIR}} = \bar{\beta}\langle\eta_k\rangle$, the initial growth rate and basic reproduction number can be reproduced by an SIR model, but a different final size $r_\infty = \sum_k r_k^\infty$ will result because

$$r_\infty = 1 - \langle\exp\left[-(\bar{\beta}/\gamma)\eta_k r_\infty\right]\rangle < 1 - \exp\left[-(\bar{\beta}/\gamma)\langle\eta_k\rangle r_\infty\right] \tag{3.35}$$

due to the concavity of the function $1 - e^{-x}$, where the average over the values of η_k is weighted by the population fractions n_k.

Case III: Heterogeneous susceptibility and infectiousness ($B_{kl} = b_k b_l$)

$$\lambda_0 = \sum_k n_k b_k^2 - \gamma \tag{3.36}$$

$$= \langle b^2 \rangle - \gamma \tag{3.37}$$

$$R_0 = 1 + \lambda_0/\gamma = \langle b^2 \rangle/\gamma \tag{3.38}$$

$$\frac{r_k^\infty}{n_k} = 1 - \exp\left\{ -\frac{b_k}{\gamma} \sum_l b_l r_l^\infty \right\} \tag{3.39}$$

By selecting the effective spreading rate $\beta_{\text{SIR}} = \langle b^2 \rangle$, the initial growth rate and basic reproduction number can be reproduced by an SIR model. Theorem 4 of [17] proves that for a given R_0, when susceptibility and infectiousness are proportional to each other in each group, the final size is less than that of a homogeneous SIR model.

Heterogeneous connectivity

In the previous section above, the rate B_{kl} at which an individual in group k infects an individual in group l can be described solely in terms of the individual characteristics of members of groups k and l. But in reality, connectivity patters tend to be clustered. In order to capture some of these clustering effects (albeit in a simplified manner), we allow for a higher probability of the disease spreading within groups than between groups. For instance, it is generally more likely for the disease to spread between two individuals living within the same city than between individuals living in different cities.

To account for this, we let $B_{kl} = b_k b_l C_{kl}$ with C_{kl} given by

$$C_{kl} = 1 - c + \frac{c}{n_k} \delta_{kl} \tag{3.40}$$

where $c \in [0, 1]$ is the clustering parameter. Note that C_{kl} is normalized such that

$$\sum_l n_l C_{kl} = 1 \tag{3.41}$$

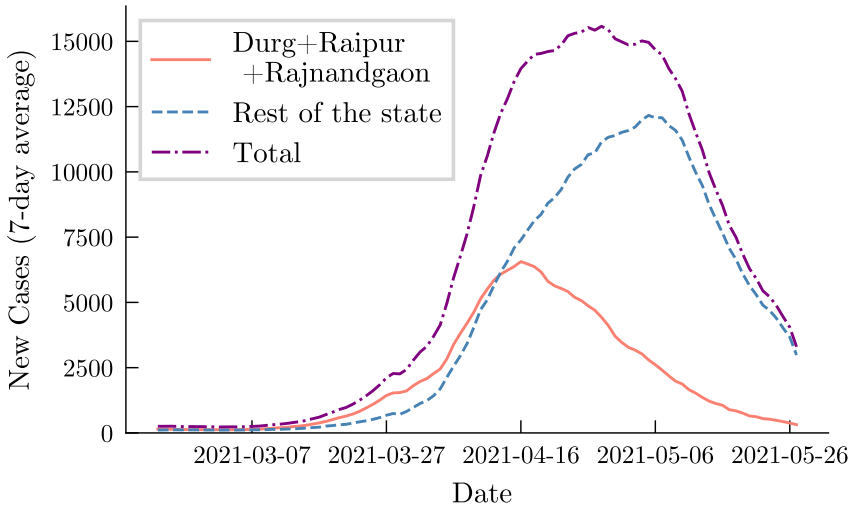


Figure 3-6: Time series of new COVID-19 infections from the state of Chhattisgarh in India shows an epidemic plateau between 2021-04-16 and 2021-05-06. Three districts of the state (out of 27), Durg, Raipur and Rajnandgaon peaked around 2021-04-16, while the rest of the districts in the state peaked around 2021-05-06. Such a disparity in the trajectories is a result of the heterogeneity in the social contact structures as well as timing and severity of the interventions. The homogeneous compartmental models when used at the scale of the state will not be able to capture the space of possible trajectories that arise out of heterogeneity. Epidemic data was obtained from [102].

In Figure 3-4 of the main text, we explore the case of two groups with different contact parameters b_1 and b_2 and a connectivity parameter c. The basic reproduction number R_0 is then given by

$$R_0 = \frac{1}{2\gamma}(n_1 B_1 + n_2 B_2 + \sqrt{(n_1 B_1 - n_2 B_2)^2 + 4 n_1 n_2 B_c^2}) \tag{3.42}$$

where $B_1 = b_1^2(1 - c + c/n_1)$, $\quad B_2 = b_2^2(1 - c + c/n_2)$, and $B_c = b_1 b_2 (1 - c)$.

Heterogeneity in the population can cause epidemic trajectories to deviate from simple trajectories (growth and decline), as shown in Figure 3-5 of the main text. One such deviation is a plateau in the epidemic trajectory observed in both simulations [234] and epidemic data. Figure 3-6 shows such a plateau in the second wave of COVID-19 epidemic in Chhattisgarh, India.

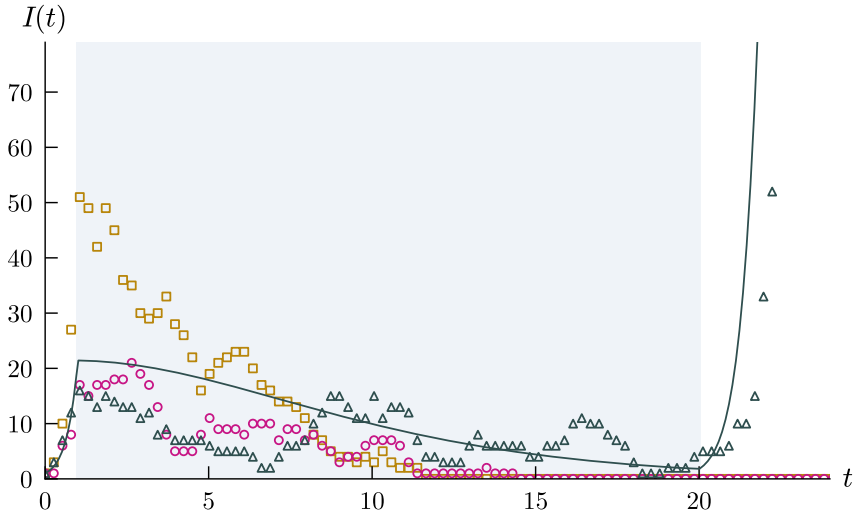


Figure 3-7: Continuous compartmental models forecast a deterministic second wave of infections. The shaded region of the plot shows the time period for which intervention is imposed by reducing the spreading rate. The blue curve shows the trajectory of number of infections in the SIR model: After interventions are removed, the infections rise again. The scatter plot trajectory shows the number of cases in the stochastic SIR model, with each marker type corresponding to a single realization of the model. The trajectories with the pink and yellow squares show that there is a finite probability for the epidemic to be eliminated during the intervention and that a second wave does not always occur. Due to the stochastic nature of disease spread, interventions cannot be held in place for a pre-determined amount of time but rather must be calibrated to real-time observations. For instance, for the case of the pink circles and yellow squares, the interventions could be lifted earlier than they were in this simulation, while for the blue triangles, the interventions would need to be kept in place for longer. Simulation parameters: $\beta = 4.1, \gamma = 1, N = 1000$. The spreading rate β is reduced by a factor of 4 during lockdown period which lasts from $t = 1.0$ to $t = 20.0$. The time step used is $\Delta t = 0.00017$.

3.5.5 Stochasticity

Elimination of an epidemic is a large-scale behavior with very a important implication: unless new cases are imported, a new wave of infections can not occur, thereby paving the way for a safe relaxation of interventions. Continuous models do not exhibit elimination behavior. Stochastic models show that if the effective reproductive number is held below 1, elimination will eventually occur (see Figure 3-7).

For Figure 3-7, we use the discrete time Markov chain (DTMC) formulation for simulating the stochastic epidemic [13]. A discrete population of size N is described by the state (S, I), where S is the number of susceptible and I is the number of infected individuals. Similar to the deterministic SIR model, β and γ are the effective spreading and recovery rates. The simulation starts with a single infected individual (the population is in a state $(N-1, 1)$), and at each successive time step the system can transition to either one of the two states: (a) a susceptible individual becomes infected, (b) an infected individual recovers; or the system can remain in the same state. The transition probabilities for each of these scenarios, when the system is in a state (S, I), are:

$$P(S-1, I+1) = \beta \frac{SI}{N} \Delta t \tag{3.43}$$

$$P(S, I-1) = \gamma I \Delta t \tag{3.44}$$

$$P(S, I) = 1 - \left(\beta \frac{S}{N} + \gamma\right) I \Delta t \tag{3.45}$$

where Δt is the time step for the simulation. Its value must be selected such that all transition probabilities lie between zero and one.

Chapter 4

Pandemic Response

Understanding what models cannot predict is sometimes more important than understanding what they can. For example, in a chaotic system such as the weather, only very short-term predictions are accurate because small changes in the present can result in very large changes in the future. The trajectory of the COVID-19 pandemic is another example: because the number of infections depends exponentially on the growth rate of the epidemic, small inaccuracies in the prediction of the growth rate will lead to large changes in the number of deaths after enough time. Furthermore, the growth or decay rate of the epidemic depends on the precise implementation details of interventions, and a very small change in the strength of interventions could be the difference between two hugely different outcomes: exponential growth until saturation versus exponential decay until elimination. Gaining an approximate understanding of the trajectory of the epidemic is important, but, when there is so much uncertainty arising from underlying disease and social dynamics in addition to the uncertainty over exactly how interventions will be implemented, detailed refinements to models are not.

More generally, spending effort trying to pin down details in models is futile if any accuracy gained is swamped by uncertainty in the measurements or by inaccuracies in the core model assumptions. What is the purpose of refining a model by 10% if there is a 50% uncertainty stemming from other aspects or assumptions of the model, or if there is a relevant behavior of the modeled system that the model fails to capture all together?

Models that attempt to capture a system's small-scale detailed behavior will inevitably include some details and leave out others. Depending on which details are included, such models may mischaracterize the system's large-scale behavior, and when they do work, it is often because their specific assumptions are a special case of a simpler, more general model. Thus, sometimes it is not the complicated models but the deceptively simple ones that are most effective for understanding a system's large-scale behavior.[1]

We now give a simple, general model that captures the effects of system details through a few larger-scale effective parameters.[2]

4.1 Introduction

The outbreak of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged in Wuhan, China, reportedly in December 2019 [354] and has since become a severe pandemic [353]. Understanding the dynamics of disease transmission both within and between regions [345, 275, 96, 98, 6, 33, 323, 34] can provide insight into how to eliminate the outbreak by imposing restrictive measures only where the virus is locally spreading, thus minimizing larger-scale economic impacts [166, 156, 23, 75]. Regions could be cities, counties, states, or any other partition of a population such that the disease transmission occurs predominantly within (as opposed to between) regions. The choice of region size depends on the spatial granularity at which policy makers are willing and able to control disease transmission.

Since it is impossible to model all the details of real-world systems, identifying large-scale behaviors is necessary to determine which details matter and how [193, 50, 298]. For COVID-19, the parameters for these large-scale behaviors include the average size of an outbreak within a region and the transmissibility of the outbreak between regions. The values of these

[1] This and the above paragraph have been taken from a preprint version of the open-access article: Alexander F. Siegenfeld, Nassim Nicholas Taleb, Yaneer Bar-Yam. What models can and cannot tell us about COVID-19. *PNAS* **117**, 16092-16095 (2020)

[2] The rest of this chapter is a slightly modified version of an article published open access under a Creative Commons Attribution 4.0 International License: Alexander F. Siegenfeld and Yaneer Bar-Yam. The impact of travel and timing in eliminating COVID-19. *Communications Physics* **3**, 204 (2020).

two parameters, both of which can be controlled with interventions, determine whether the behavior of the epidemic within a collection of regions (e.g. a country) is that of exponential spread until saturation (e.g. dynamic endemic equilibrium or herd immunity) or exponential decay until elimination. In the latter regime, elimination will be stable and most regions within the collection can fully open up their economies, with only local and sporadic social distancing measures needed to contain outbreaks arising from hidden or imported cases.

A central concept in the study of disease spread is the reproductive number R, i.e. the average number of people to whom a typical infected individual will transmit the disease [119]. To consider the transmission between regions, an analogous *region-to-region reproductive number* R_* can be defined as the number of other regions (including those that have been previously infected) to which a single infected region will transmit the infection on average [35, 97]. Just as an outbreak cannot sustainably propagate (i.e. elimination is stable) among individuals within a region if $R < 1$, an outbreak cannot sustainably propagate *among regions* within a collection if $R_* < 1$.

Here, we analyze how local measures can support the elimination of COVID-19 while avoiding large-scale lockdowns where they are unnecessary. We find that reductions in travel and the speed with which regions act to contain future outbreaks play decisive roles in whether COVID-19 is eliminated from a collection of regions (i.e. whether $R_* < 1$). Such an elimination is stable: for each outbreak caused by imported or undetected cases, only a few regions—fewer than $1/(1 - R_*)$ on average—need to temporarily impose measures while the rest of the regions in the collection remain open. If infected regions (including those that become re-infected in the future) impose control measures shortly after experiencing community transmission (i.e. shortly after no longer being a "green zone"), the number of infected regions, and thus the number of regions in which such measures are required, will exponentially decrease over time.

4.2 Results

4.2.1 General model

The disease is modeled as being transmitted among individuals within a region, with travel allowing the disease to spread between regions. A collection of regions is defined as any partition of a population such that travel/social contact within each region far exceeds that between them (e.g. the U.S. could be partitioned by county or commuting zone boundaries), in order that an infected individual is far more likely to transmit the virus to someone within the same region rather than to someone in a different region (i.e. transmission between regions can be treated as a first-order perturbation). Policy makers are assumed to act homogeneously within each region and to have the ability to act independently between regions—thus, the choice of how to partition a population into regions must take into consideration the spatial granularity at which policy makers are willing and able to make decisions. Treating larger areas as single regions means that social distancing measures will be homogeneously applied to larger areas but also means that it is easier to achieve lower per capita travel rates between such areas.

We define a region as *infected* if someone with the infection enters the region. Community transmission, also known as community spread, is said to occur when individuals within a region are infected from an unknown source [85]. Conditioning on region c being infected, we let N_c be a random variable (that could be zero) denoting the total number of new infections that occur in region c from the time after the region is infected to the time at which there is no more community transmission, at which point we define the region as no longer infected. Let p_c be the probability that an infected individual in region c will travel to another region during the period in which that individual is contagious. Then, the region-to-region reproductive number for region c—which we define as the expectation of the number of regions that region c will infect if it becomes infected—is $R_*^c = \mathbb{E}[N_c]p_c$. The expected outbreak size $\mathbb{E}[N_c]$ can be greatly reduced if regions impose strong social distancing measures shortly after detecting community transmission (these social distancing

measures can be lifted once community transmission has been contained), while p_c can be reduced by reducing travel out of infected regions.

R_*^c may differ from region to region, with R_*^c for a particular region depending not only on internal factors but also on the network connectivity between that region and others, which in turn depends on the sizes of regions (i.e. the scale/level of granularity of the interventions). If the interventions are sufficiently fast and strong such that R_*—the average value of R_*^c over a collection of regions with each region weighted by its probability of being infected [119]—is less than one, then the outbreak will not be self-sustaining within that collection of regions. Put another way, a collection of regions can exist in one of two regimes: a regime for which elimination is a stable fixed point of the system and a regime for which it is unstable (see fig. 4-1). A change in policy can shift a collection of regions from the unstable regime ($R_* > 1$) to the stable regime ($R_* < 1$) or vice versa. Although the values of N_c for a collection of regions could currently be high, R_* is determined by $\mathbb{E}[N_c]$ and p_c for the regions that will be infected or re-infected in the future after social distancing has eliminated the virus from currently infected regions.

We note that if region c were infected multiple times, $\mathbb{E}[N_c]$ would be higher than if it were infected once, but it is assumed that infecting an already infected region will on average contribute no more to disease spread than infecting a currently uninfected region. Thus, like the basic reproductive number, this region-to-region reproductive number overestimates the disease spread away from the limit of most regions being uninfected by counting a single region that has been infected multiple times during a single outbreak as multiple regions being infected. To the extent that a region being infected multiple times has a linear effect on its expected total number of cases, this approximation will not impact the values of R_*. However, if a region that is currently implementing control measures because of previous importations receives additional imported infections, these additional importations are not likely to have as large an effect as the previous ones that occurred before control measures were implemented. Thus, R_* may be overestimated when a high total number of infections within a collection of regions makes it common for regions to experience multiple

importations during a single outbreak. We also note that, as in SIS compartmental models, an infected region in which the virus is contained can later be re-infected.

It is important to keep in mind that a collection of regions can be in the stable regime, in which the *region-to-region* reproductive number for the collection R_* is less than one, while, at any given time, the reproductive number R of most regions within the collection is greater than one. In other words, in order for $R_* < 1$ for the collection of regions, individual regions within the collections need only maintain $R < 1$ so long as community transmission persists within them; otherwise, they can lift social distancing measures and open up economically. For every infection that a collection of regions with $R_* < 1$ imports (not including cases that were quarantined at the border), the average number of regions within that collection that will need to impose social distancing measures is bounded by $\alpha \sum_{n=0}^{\infty} (R_*)^n = \alpha/(1 - R_*)$ where α, which can be reduced with testing and contact tracing, is the probability that an imported case will result in community transmission within a region. Thus, border control need not be perfect; if a collection of regions has sufficiently good border control and policies that ensure $R_* < 1$, elimination will remain stable while most of the collection's regions remain open most of the time.

4.2.2 Modeling the approximate size of regional outbreaks

We now describe a simple mathematical model to estimate N_c, not to provide a precise description of the epidemic trajectory but rather to clarify how various interventions may affect outbreak size. This specific model assumes exponential growth before control measures are implemented followed by exponential decline afterwards. However, the validity of this assumption is not essential to the main results, which depend only on the number of infected individuals traveling out of each region. Modeling the number of active cases within a region using exponential growth and decay serves mainly to give a quantitative handle on the rates of increase and decrease in cases, but comparable results could be obtained using other epidemic trajectories. (Similarly, although SIS and SIR models assume an exponential distribution of generation intervals, this particular assumption does not affect the conditions

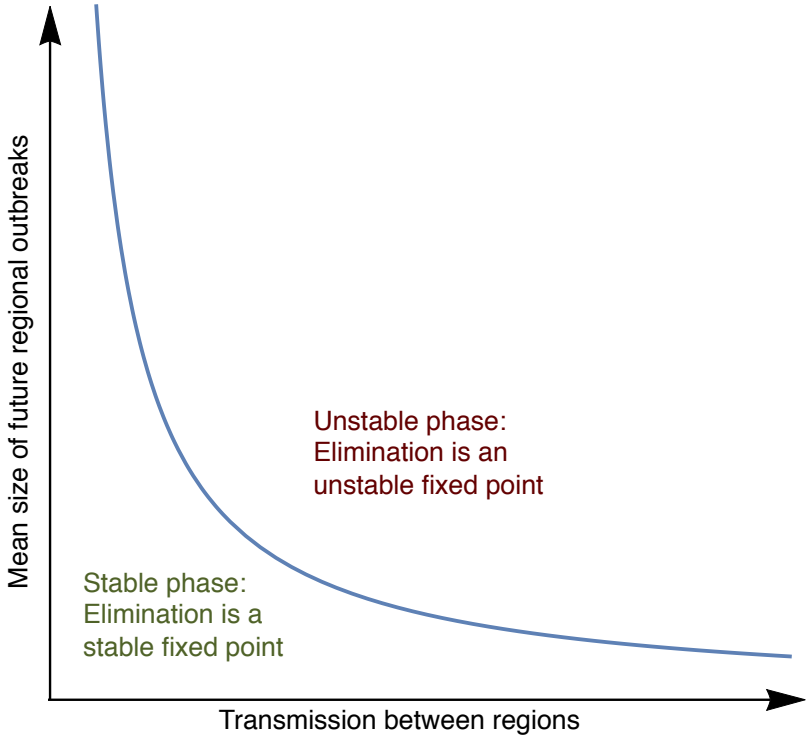


Figure 4-1: **Phase diagram of COVID-19 for a collection of regions.** Elimination can either be stable or unstable; the stability of elimination is a function of (1) the average total number of cases that will result from the disease being transmitted to a region (which depends on, among other factors, how quickly regions locally impose social distancing measures if they are infected or re-infected), and (2) the probability that an infected individual in one region will infect an individual in another (which depends on the rate of travel between regions). Note that the stable regime does not require that every region implement social distancing measures but rather only those with active community transmission. Thus, once elimination is achieved, it can be maintained while most regions remain economically open, with outbreaks caused by hidden or imported cases contained by social distancing measures that are localized in both space and time. Image source: [300].

under which such models are valid, so long as their recovery rate γ is treated as an effective parameter.) The one key assumption is that the reproductive number within individual regions can be reduced below one with sufficiently strong measures such as mask-wearing, social distancing, contact tracing, etc. This assumption has been validated empirically by the many regions around the world in which the number of cases has declined for a sustained period of time.

Let i_0^c be a stochastic factor that roughly corresponds to the initial foothold that the virus gains in region c conditioning on an infected individual entering the region, with $i_0^c = 0$ corresponding to the case in which no one was infected or a few people were infected but the outbreak was contained by contact tracing/quarantine or otherwise spread no further (e.g. no community transmission). The probability distribution of i_0^c will depend on the distribution of infectiousness [224], a function of both biological and social factors. If the outbreak is contained ($i_0^c = 0$), then the number of active infections remains at zero for the purposes of this model because—by definition—the outbreak has no chance of spreading to other regions.

If the outbreak is not contained ($i_0^c \neq 0$), the number of active infections is modeled as growing with time t at an exponential rate $e^{r_c t}$. Then, after time T_c (the delay in response), the region implements social distancing measures that cause the number of active infections to decay as e^{-t/τ_c}, where τ_c is the time-constant (i.e. the inverse of the rate) of decay. Such exponential decrease will occur if the social distancing measures, together with mask-wearing and testing/contact tracing/quarantine, can reduce the reproductive number (R) of the virus below one. The greater the reduction in R, the smaller the value of τ_c and thus the faster the decrease in infections. (For $R < 1$, R is related to τ_c by $1 = \int_0^\infty R g(t) e^{t/\tau_c} dt$ where $g(t)$ is the distribution of generation intervals [344].) We note that this assumption of exponential increase followed by exponential decay after intervention assumes that the proportion of susceptible individuals is roughly constant, i.e. that the region intervenes before a significant fraction of its population is infected, which is the regime with which we are concerned. To the extent that this assumption does not hold, its use will result in

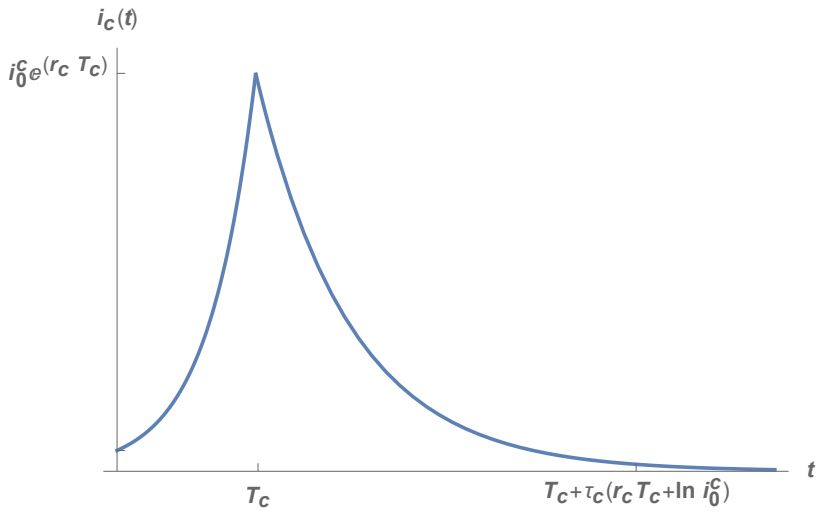


Figure 4-2: **Representative epidemic trajectory of an outbreak within a single region.** The number of active infections in a region c according to the model described by eq. (4.1) are shown as a function of time. The time at which community transmission begins is defined as $t = 0$. At $t = T_c$ (a duration of T_c after the start of community transmission), control measures that reduce the reproductive number of the virus to less than one are implemented. At $t = T_c$, the size of the outbreak is $i_0^c e^{r_c T_c}$ where i_0^c is a stochastic factor and r_c is the rate of exponential growth without the control measures. In order to eliminate the community transmission, we estimate that the control measures must remain in place for a duration of approximately $\tau_c(r_c T_c + \ln i_0^c)$ (see eq. (4.2)), where τ_c^{-1} is the rate of exponential decay in the number of infections under the control measures. Thus, the longer the region waits to enact the measures, the longer the total amount of time they must remain in place.

an overestimate of the number of infected individuals and thus does not affect our main conclusions.

The number of active infections in region c as a function of time (see fig. 4-2) can therefore be written as:

$$i_c(t) = \begin{cases} i_0^c e^{r_c t} & t \leq T_c \\ i_0^c e^{r_c T_c} e^{-(t-T_c)/\tau_c} & t \geq T_c \end{cases} \tag{4.1}$$

The social distancing measures can be lifted once no active infections remain in the region or once there is no more community transmission and the remaining infections can be contained with contact tracing. Solving for $i_c(t) = 1$ (assuming $i_0^c \neq 0$) yields an approximate

duration for the social distancing measures of

$$\tau_c(r_c T_c + \ln i_0^c) \tag{4.2}$$

As the number of cases becomes increasingly small, testing and contact tracing become increasingly effective and can hasten the end of community transmission, thereby allowing the social distancing measures to be lifted. The probability that the number of infections will rebound after the social distancing measures are lifted—in which case an additional phase of such measures will be needed, as in the Imperial College report [137]—will depend on the probability of importation from other regions, which in turn will depend on the region-to-region reproductive number R_*. Even though the virus can be re-imported, as long as $R_* < 1$, the number of infected regions will on average decrease over time, since re-importation events are included in R_*.

4.2.3 Modeling transmission between regions

Each infected region c infects a currently uninfected region with a probability rate proportional to the number of active infections $i_c(t)$ times the probability rate p_c that an infected individual will travel to an uninfected region. The number of new infected regions spawned by region c can thus be modeled as a Poisson process with rate $i_c(t)p_c$. As described above, this modeling assumption overestimates the spread of the disease to new regions by counting a single new region that has been infected multiple times (possibly by multiple other regions) as multiple new infected regions. The smaller the number of regions infected and the smaller the probability of one region infecting another, the smaller the probability that the same region will be infected twice. Nonetheless, for certain regional connectivity networks, this model may overestimate R_*. (Our main conclusions are unaffected because R_* will be less than one if its overestimate is.)

Let p_0^c be the per capita probability rate before time T_c of individuals in region c traveling to other regions and p_1^c be the probability rate afterwards (p_1^c will be less than p_0^c if travel is

discouraged and/or restricted at the time social distancing measures are implemented). The number of new regions that are infected by region c will then be a Poisson random variable with a mean of

$$i_0^c p_0^c \int_0^{T_c} e^{r_c t} dt + i_0^c p_1^c e^{r_c T_c} \int_0^{\tau_c (r_c T_c + \ln i_0^c)} e^{-t/\tau_c} dt \tag{4.3}$$

$$= \begin{cases} i_0^c \left(p_0^c \frac{e^{r_c T_c} - 1}{r_c} + p_1^c \tau_c (e^{r_c T_c} - \frac{1}{i_0^c}) \right) & i_0^c > 0 \\ 0 & i_0^c = 0 \end{cases} \tag{4.4}$$

Taking the expected value over i_0^c (and allowing for a slight overestimate of R_*^c by treating the $\frac{1}{i_0^c}$ term as negligible when $i_0^c > 0$) yields

$$R_*^c = \mathbb{E}[N_c] p_c = \mathbb{E}[i_0^c] \left(p_0^c \frac{e^{r_c T_c} - 1}{r_c} + p_1^c \tau_c e^{r_c T_c} \right) \tag{4.5}$$

(The value of R_* for a collection of regions is then an appropriately weighted average of R_*^c over that collection, as described above.)

4.2.4 Parameter estimation for COVID-19

In order to better understand the extent of the measures required to achieve $R_* < 1$, we estimate the values of the parameters in eq. (4.5) (see fig. 4-3 and Methods) in order to determine R_*^c as a function of the time-delay before social distancing measures are enacted, as shown in fig. 4-4. Note that the time-delay is measured from the time at which exponential growth begins to occur—which could be as early as the beginning of community transmission within the region—not the time at which exponential growth is first measured. The latter may lag the former due to delays in testing and the possibility of pre-symptomatic/asymptomatic spread.

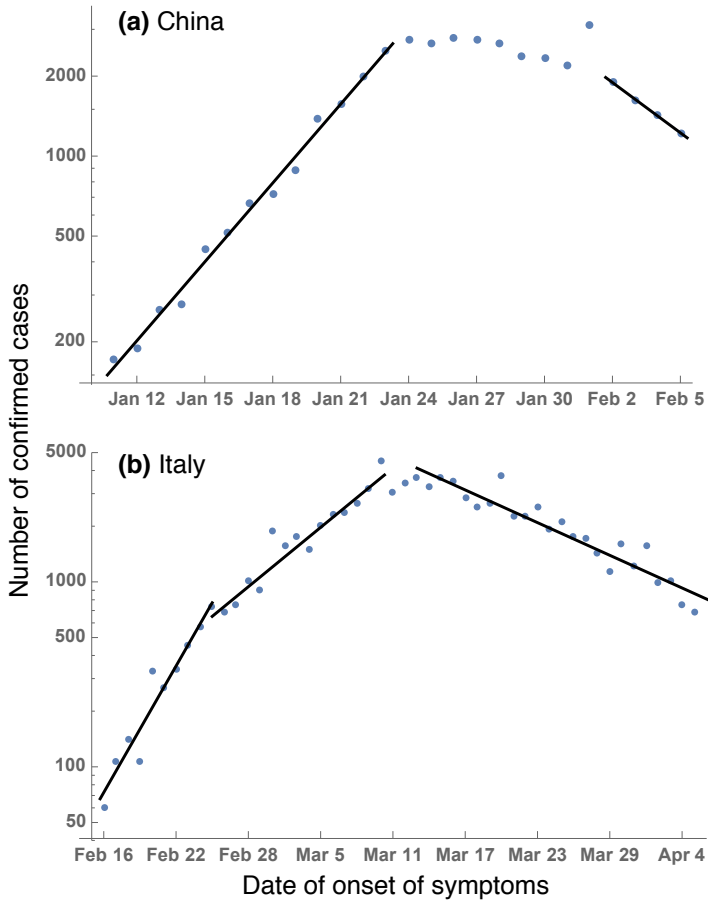


Figure 4-3: **Logarithmic plots of confirmed cases by date of symptom onset in China and Italy.** **(a)** The daily number of confirmed cases in China—by date that these patients self-reported as the onset of their symptoms—are shown as dots on a logarithmic scale. The solid lines are the best ordinary least squares linear fits to the natural logarithm of the number of cases: For Jan. 11-23 (up until the lockdown), the slope (in units of day^{-1}) is 0.228 ($R^2 = 0.991$, 95% confidence interval (CI) [0.214, 0.242]), which corresponds to a doubling time of 3.04 days. For Feb. 2-5 the slope is -0.145 ($R^2 = 0.999$, 95% CI [-0.160, -0.131]), which corresponds to a halving time of 4.78 days. Data are from the Chinese Center for Disease Control and Prevention [359], which includes cases diagnosed through Feb. 11. Not pictured: There is a drop in cases with onsets of symptoms after Feb. 5 [359], likely due to many of those cases being diagnosed after Feb. 11. **(b)** The daily number of confirmed cases in Italy by date of symptom onset are shown as dots on a logarithmic scale (data are from Italian authorities [184]). The best ordinary least squares linear fits are shown as solid lines and have slopes (in units of day^{-1}) of 0.262 ($R^2 = 0.927$, 95% CI [0.202, 0.322]) for Feb. 16-25, 0.123 ($R^2 = 0.923$, 95% CI [0.102, 0.144]) for Feb. 25 - Mar. 10, and -0.068 ($R^2 = 0.901$, 95% CI [-0.078, -0.058]) for Mar. 13 - Apr. 5. The change in the exponential growth rate from 0.262 to 0.123 likely occurred due to partial measures implemented by Italy, but it was not until a nationwide lockdown was implemented on March 9 that exponential growth changed to exponential decline. The rate of decline is much slower in Italy than in China, perhaps due to China's stronger lockdown enforcement and contact tracing/quarantine measures.

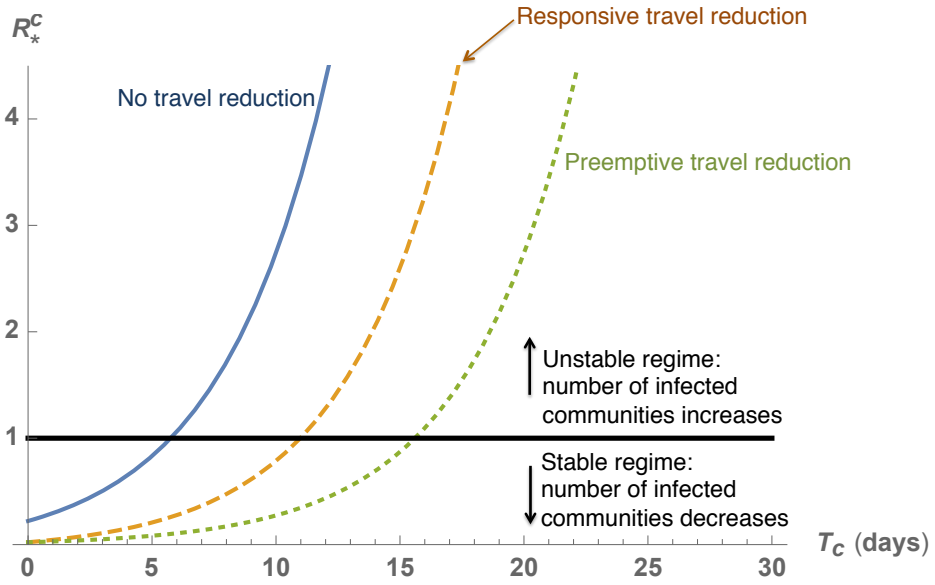


Figure 4-4: **Dependence of the region-to-region reproductive number on travel policies and the time delay before regions that become infected impose social distancing measures.** Each curve depicts, under different travel policies, the average number of regions R_*^c to which region c will transmit the disease as a function of the time delay T_c between when region c experiences community transmission and when it imposes social distancing measures (eq. (4.5)). If the region-to-region reproductive number R_* (a weighted average of R_*^c over a collection of regions) is less than one, the number of infected regions will exponentially decrease and the elimination of the disease within that collection of regions will be stable; otherwise, the number of infected regions will increase until saturation. **Parameter values:** All curves use $\mathbb{E}[i_0^c] = 3.7$, $\tau_c = 15$ days, and $r_c = 0.228$ day^{-1}. Solid curve: no travel reduction; $p_0^c = p_1^c = 0.004$ day^{-1}. Dashed curve: 10-fold (responsive) travel reduction after time T_c; $p_0^c = 0.004$ day^{-1} and $p_1^c = 0.0004$ day^{-1}. Dotted curve: general (preemptive) 10-fold travel reduction; $p_0^c = p_1^c = 0.0004$ day^{-1}. Note that some of these parameter estimates are conservative and are likely to overestimate R_*^c as a function of T_c—see Methods for discussion.

4.3 Discussion

Our analysis and parameter estimates are conservative and are therefore likely to overestimate R_*, meaning that there may be more room for error than fig. 4-4 suggests. However, given the considerable uncertainty surrounding pandemics and the impossibility of precisely predicting their trajectories [300], R_* should be reduced as much as possible so as not to take any chances, as well as so that the virus will be eliminated (and economies more fully reopen) as quickly as possible. Since R_* is proportional to both $\mathbb{E}[N_c]$ and the travel rates between regions, any intervention that reduces the sizes of outbreaks within regions and/or travel between them will also reduce R_*. In the language of eq. (4.5) (see table 4.1):

- A reduction in travel from region c results in a linear reduction in R_*^c through p_0^c and p_1^c.

- Improvements in testing, contact tracing, and quarantine reduce R_*^c through $\mathbb{E}[i_0^c]$, r_c, and τ_c.

- Preemptive measures (i.e. measures before T_c, including when no spreading has been detected) such as mask-wearing and the reduction of large gatherings reduces the probability of a super-spreader event as well as general transmission, reducing R_*^c through both $\mathbb{E}[i_0^c]$ and r_c. Because—in the absence of interventions—a small fraction of COVID-19 cases cause most of the spread [278, 130], the reduction of super-spreader events can have an outsized impact.

- Reductions in $r_c T_c$ not only exponentially reduce R_*^c (as well as the total number of infections within the region) but also linearly reduce the amount of time for which the social distancing measures must remain in place. Early-detection systems [3, 200] and more comprehensive testing may greatly reduce T_c.

- Stronger social distancing measures (after time T_c) decrease τ_c, which results in a linear decrease in both R_*^c and the time for which the distancing measures must remain in place.

Table 4.1: Summary of parameters affecting the region-to-region reproductive number

Param.	Description
r_c	The rate of exponential growth in active infections in region c before control measures are implemented.
τ_c	The time-constant (inverse rate) of exponential decline in active infections after the control measures are implemented.
T_c	The time delay between the start of exponential growth and the implementation of the control measures.
i_0^c	A stochastic measure of the initial foothold that the virus gains prior to exponential growth, with $i_0^c = 0$ corresponding to containment.
p_0^c	The probability rate that an infected individual in region c will travel outside the region before the control measures are imposed.
p_1^c	The probability rate that an infected individual in region c will travel outside the region after the control measures are imposed.

We conclude with a few comments. First, without the timely implementation of control measures that reduce the local reproductive number to less than one, restricting travel from infected regions serves only to delay the spread of the outbreak, as found in other studies [27, 270, 74, 147, 133, 349, 89]. However, when a reduction in travel is coupled with such measures, the travel reduction will not only delay the spread of the outbreak but in some cases will also be the determining factor in whether or not the outbreak is eliminated. (If $R_* < 1$ can be achieved without reducing travel, travel reductions can greatly decrease the duration and total case count of the epidemic by further reducing R_*.) Empirically, travel restrictions, when combined with other sufficiently strong interventions, have been found to substantially curb the COVID-19 epidemic [203].

Second, because R_*^c depends exponentially on T_c, the longer a region waits to implement social distancing measures, the more important it becomes to act without delay. It is important to note, however, that there is no advantage to delaying for even a short time. Immediately implementing social distancing measures as soon as there is evidence of the disease spreading within the region will not only reduce the total amount of time for which such measures must remain in place but will also exponentially reduce the probability of infecting or re-infecting another region. Thus, there is no tradeoff here between health and economics—acting quickly will reduce the duration of both the disease and the economic harm.

Practically speaking, all regions within which the virus is spreading must implement and maintain control measures that reduce the local reproductive number to less than one until the virus has been eliminated or contained. If a locale experiences another outbreak of community transmission (e.g. because the virus was re-imported), sufficiently strong interventions should be implemented in a large enough region around that locale so that the per capita rate of travel out of the region is sufficiently low. Per capita travel rates between regions do not have to be zero, but they must be small—precisely how small depends on the parameters in eq. (4.5). In the event that some regions "defect" by not containing their outbreaks (thereby putting other regions at risk), other regions must either quarantine trav-

elers from those regions or maintain precautions that keep their local reproductive numbers below one to avoid further outbreaks. Due to the higher rates of travel between neighboring regions, extra precautions are necessary in a region that has a neighbor with widespread disease transmission.

Finally, like transmission within a region, transmission between regions is an exponential process. At first the number of infected regions is deceptively small, but if $R_* > 1$, this number exponentially grows. If, however, a collection of regions adopts a set of protocols that achieves $R_* < 1$, sustained propagation of the disease between the regions will not be possible, and future outbreaks caused by importations will die out while leaving most regions within the collection unaffected.

4.4 Methods

The doubling time of the epidemic can vary from location to location and depends on pre-lockdown interventions (see e.g. the change in the growth rate for Italy in fig. 4-3). Using the number of confirmed cases in China by date of symptom onset (rather than by date of diagnosis) [359] yields a doubling time of 3.04 days in the period leading up to the Jan. 23 lockdown, which corresponds to $r_c = 0.228$ day^{-1} (fig. 4-3). Some studies estimated the doubling time for COVID-19 at approximately 7 days [357, 216], but even a 5 day doubling period is implausibly long, given that in various countries, even with some preventative measures, the number of infections increased by far more than a factor of 64 over a 30 day period [355]. Part of the difficulty in estimating the doubling time from the initial period of transmission is that 'super-spreader' events may play an important role in the transmission process [224, 278, 130]. The presence of super-spreader events indicates that the transmission process may be dominated by relatively uncommon events and that therefore standard statistical approaches may underestimate the rate of spread—and thus overestimate the doubling time—when the total number of cases is still small [320]. (Overestimates of the doubling time lead to underestimates of R_0.) Furthermore, due to heterogeneity from both social and biological factors in transmission rates among various subgroups in a population,

the expected growth rate in the very earliest phase of an epidemic depends on the subgroups in which the disease is initially introduced and may not be reflective of the true reproductive number [119].

While $i_c(t)$ is equal to the number of active infections rather than the daily number of new cases, the doubling time for each will be equal during a period of exponential growth, and therefore the exponential rate of growth of the daily number of new cases can be used to estimate r_c. We note that a doubling time of 3 days will likely be a substantial overestimate for regions that, even when cleared of the virus, still maintain precautionary measures such as mask-wearing, working from home when possible, and avoiding large gatherings.

From this growth rate $r = 0.228$ day^{-1} in China before the Jan. 23 lockdown, the basic reproductive number R_0 (which also varies by location and time) can be calculated for China before Jan. 23 using

$$\frac{1}{R_0} = \int_{-\infty}^{\infty} g(t)e^{-rt}dt \tag{4.6}$$

where $g(t)$ is the distribution of generation intervals [344]. Empirically, we generally observe the distribution of serial intervals (the times between the onsets of symptoms in two successive cases in a transmission chain) rather than the distribution of generation intervals (the times between two successive infections in a transmission chain) [261]. The means of the two distributions will, however, be the same. The distribution of generation intervals is affected by non-pharmaceutical interventions; for example, if infected individuals are quarantined, the transmission that is not prevented will more likely occur at shorter generation intervals. In the period before the Jan. 23 lockdown, the mean serial interval (which equals the mean generation interval) was estimated to be 7.8 days, significantly longer than estimates of 5.1 days and 2.6 days for the time periods Jan. 23-29 and Jan. 30-Feb. 13, respectively [11]. As we are estimating R_0 in China before Jan. 23, we focus on the distribution of generation intervals from that period (with the mean of 7.8 days). If all transmission occurred at the estimated mean generation interval of 7.8 days, $R_0 = e^{7.8r} = 5.9$. However, due to the spread of generation intervals, $R_0 = 5.9$ will be a substantial overestimate. R_0 can be underestimated by approximating the distribution of generation intervals $g(t)$ by the distribution of serial

intervals (estimated as a normal distribution with mean 7.8 days and standard deviation 5.2 days [11]), which yields $R_0 = 2.9$. ($R_0 = 2.9$ is an underestimate because the distribution of serial intervals generally has a larger variance than the distribution of generation intervals, and serial intervals, unlike generation intervals, can be negative [261].) Thus, for China before Jan. 23, $2.9 < R_0 < 5.9$, but without a distribution of generation intervals it is hard to be significantly more accurate. We estimate $R_0 = 3.7$, which can be obtained by taking the distribution of generation intervals $g(t)$ to be uniformly distributed between 0 and 15.6 days (15.6 being chosen so that the mean generation interval is 7.8 days).

The values of τ_c that can be achieved depend on the effectiveness of the social distancing measures. The data from China (see fig. 4-3) indicate a halving time of as few as 4.78 days is achievable, which corresponds to $\tau_c = 6.9$ days. However, as a more conservative estimate, we use $\tau_c = 15$ days from Italy's data, which exhibited a particularly slow decline in cases compared with most other European countries [355]. (The halving time of the daily number of new cases will equal the having time of the number of active infections $i_c(t)$, which is why τ_c can be estimated from this data.)

$\mathbb{E}[i_0^c]$ is the expected "effective" number of people an infected traveler will infect while visiting region c, taking into account containment efforts. For instance, if the outbreak is contained such that exponential growth never occurs, the effective number of people infected by the traveler is zero, even if the traveler did infect some individuals in region c. We estimate that in the absence of any mitigating policies (mask-wearing, testing, etc.), $\mathbb{E}[i_0^c] \sim R_0 \sim 3.7$; the degree to which $\mathbb{E}[i_c^0]$ differs from R_0 depends on how likely a typical traveler is to transmit the virus relative to a typical resident, as well as on the effectiveness of contact tracing and other containment efforts. For regions in which preemptive measures (e.g. mask-wearing, avoiding large gatherings, etc.) are taken, we expect that $\mathbb{E}[i_0^c] \sim R_0 \sim 3.7$ is a substantial overestimate since there is a far greater chance that an infected traveler will not spark an uncontained outbreak (and thus i_0^c will be 0). Good testing and contact tracing policies can also substantially increase the chance that $i_0^c = 0$ (thereby reducing $\mathbb{E}[i_0^c]$) by reducing the probability of community transmission. And a rigorous enough quarantine policy at the

border of a region may reduce $\mathbb{E}[i_0^c]$ to nearly zero by preventing infected travelers from even having a chance to infect individuals in the region.

The value of p_c depends on the frequency of travel out of region c. As previously noted, there is some choice in how to partition a population into a collection of regions. In general, the larger the regions, the lower the frequency of per capita travel out of them but the more homogeneous the application of the social distancing measures. Since p_c is smaller for larger regions and N_c is not strongly affected by the size of regions, R_* will be lower if larger regions are chosen, but at the cost of the social distancing measures being applied over larger areas for each new outbreak.

Regions do have to be large enough so that transmission between regions can be treated as a first-order (linear) perturbation to a system in which most spread occurs within regions. Thus, how small the regions can be depends on the extent to which travel between them can be reduced. A region could be as small as, for example, a neighborhood if the neighborhood was willing to take measures so as to greatly reduce contact with people outside of it. (We note that the regions within any given collection can differ in both geographic size and in population from one another.)

Although p_c will depend on the size of region c as well as the travel behavior of individuals within that region, we still wish to get an estimate of a plausible value for p_c. Considering a collection of regions within the U.S. that are large enough such that travel between the regions is predominantly by flight yields a per capita travel rate of 0.004 flights out of a region per person per day. This estimate is obtained by dividing the 1.01 billion total passengers traveling by plane to, from, or within the U.S. in 2018 [337] by the 2018 U.S. population and the number of days in 2018, and then also dividing by 2 so that only flights out of and not into a region are counted. Using this estimate for p_0^c assumes that the probability that an infected individual will travel equals that of the general public. Travelers may on average have more social contacts than the general public or may become infected because of their travels, which could mean that infected individuals are more likely to be traveling than the general public. On the other hand, an individual who has COVID-19 symptoms

or has tested positive for COVID-19 may be less likely to travel than the general public. It should also be noted that, compared to 2019 numbers, U.S. air travel in April and May of 2020 was down by approximately a factor of 10 (the effects of a 10-fold reduction in travel are shown in fig. 4-4) and U.S. air travel in June was down by approximately a factor of 5 [336].

Chapter 5

Unmasking the Mask Studies: The Importance of Theory

There is no such thing as purely empirical observations; all data require a framework in which they can be interpreted, and assumptions must always be drawn in order to link data to conclusions in scientific studies. However, perhaps due to a lack of ability to distinguish between good and bad theory, many fields have an implicit distrust of theory and modeling in general (see section 3.1). The danger is that a distrust of theory doesn't result in no theory being used; rather, it leads to the use of a default theory in which the assumptions are not made explicit. The default theories used when considering empirical studies tend to assume independence: for instance, if an empirical study—even a randomized controlled trial—shows that X causes Y, such a result is meaningful only to the extent the effect of X on Y is independent of the context of the trial, heterogeneity within the sample of individuals, etc. Such assumptions of independence, especially when they remain implicit, can result in misleading conclusions for complex systems, which are precisely those for which various causes are strongly interconnected.

In this chapter, we give an account of how a neglect of theory resulted in an incorrect interpretation of the evidence regarding the efficacy of surgical masks at the beginning of

the COVID-19 pandemic.[1] The error was twofold. First, no explicit assumptions were made relating the frequency of mask use to the probability of infection. While concerns about the frequency of mask use in the studies were often qualitatively raised, such theory was not quantitatively included in previous meta-analyses [188]. In other words, there was an implicit assumption within the mathematics of the meta-analyses that the efficacy of masks is independent of how often they are worn. As shown below, a relatively simple model linking mask use frequency to infection probability can correct this error; indeed, good theory is often simple (see also chapter 4), and overly complicated theory could be part of the reason why it is distrusted in the first place.

Second, naive empiricism was trusted over the simple theoretical consideration that given what we know about how respiratory viruses are transmitted, and given what we know about how masks work, we can conclude that masks will provide at least some protection against respiratory viruses, a conclusion that was immediately drawn by countries such as Taiwan and Japan. Interpretations of data can be flawed in many ways, and there is also a legibility bias: searching only for conclusions that can be drawn from hard data is analogous to the drunk man searching for his keys under the streetlamp.

5.1 Introduction

In 1910, one of the first western-trained Chinese physicians adapted surgical masks for use against a respiratory plague that killed more than 60,000 people in four months [155]. The logic behind their function is transparent: a mask can block some viral or bacterial particles from entering and/or dispersing from the wearer's respiratory tract. They have been used for prevention in a wide range of disease outbreaks and medical settings, and there is currently a general consensus that surgical and cloth masks help prevent infected individuals from spreading COVID-19. Surprisingly, given the logic of their utility, there is less of a consensus

[1]Most of this chapter is a slightly modified version of the preprint `https://arxiv.org/abs/2102.04882`, which was later published as: Pratyush K. Kollepara, Alexander F. Siegenfeld, Nassim Nicholas Taleb, Yaneer Bar-Yam. Unmasking the mask studies: why the effectiveness of surgical masks in preventing respiratory infections has been underestimated. *Journal of Travel Medicine* **28**, taab144 (2021).

that surgical/cloth masks also protect the wearer and many government health organizations did not initially recommend wearing them during the early months of the COVID-19 pandemic.

It is well established that surgical and cloth masks partially block virus-containing airborne droplets of various sizes [64, 143, 221, 250, 14, 339, 215, 109, 201, 94]. Cloth masks, surgical masks, respirator masks (e.g. N95), and powered air-purifying respirators are understood to be capable of providing increasing levels of protection. The amount of virus transmitted between an infected and a susceptible individual is therefore expected to be reduced if either is wearing a mask, with both wearing masks giving the best protection. However, this straightforward inference has been difficult to establish in experimental studies. Here we analyze why some experimental studies find masks to be effective while others do not. We determined that the studies that did not find surgical masks to be effective were under-powered to such an extent that even if the masks were 100% effective, they still would have been unlikely to find a statistically significant result. Statistical power is the probability that a study will find a statistically significant result if its intervention does in fact have a certain effect. Our results concerning the statistical power of mask studies are summarized in fig. 5-1, which shows that all studies that had a large enough sample size and/or adherence for 80% power (above and to the right of the gray lines) show a statistically significant reduction in infections among mask-wearers. As would be expected, most studies with less statistical power (towards the lower left) did not find a statistically significant effect. We also provide a framework for understanding the nonlinear effects of mask-wearing on the probability of infection. Experiments that do not take such factors into account provide misleading results unless interpreted carefully. While the precautionary principle [159] would recommend the use of masks during the COVID-19 pandemic in any case (due to the asymmetric risks of using vs. not using masks), the analyses we provide gives consistency to theoretical analyses, experimental studies, and epidemiological recommendations.

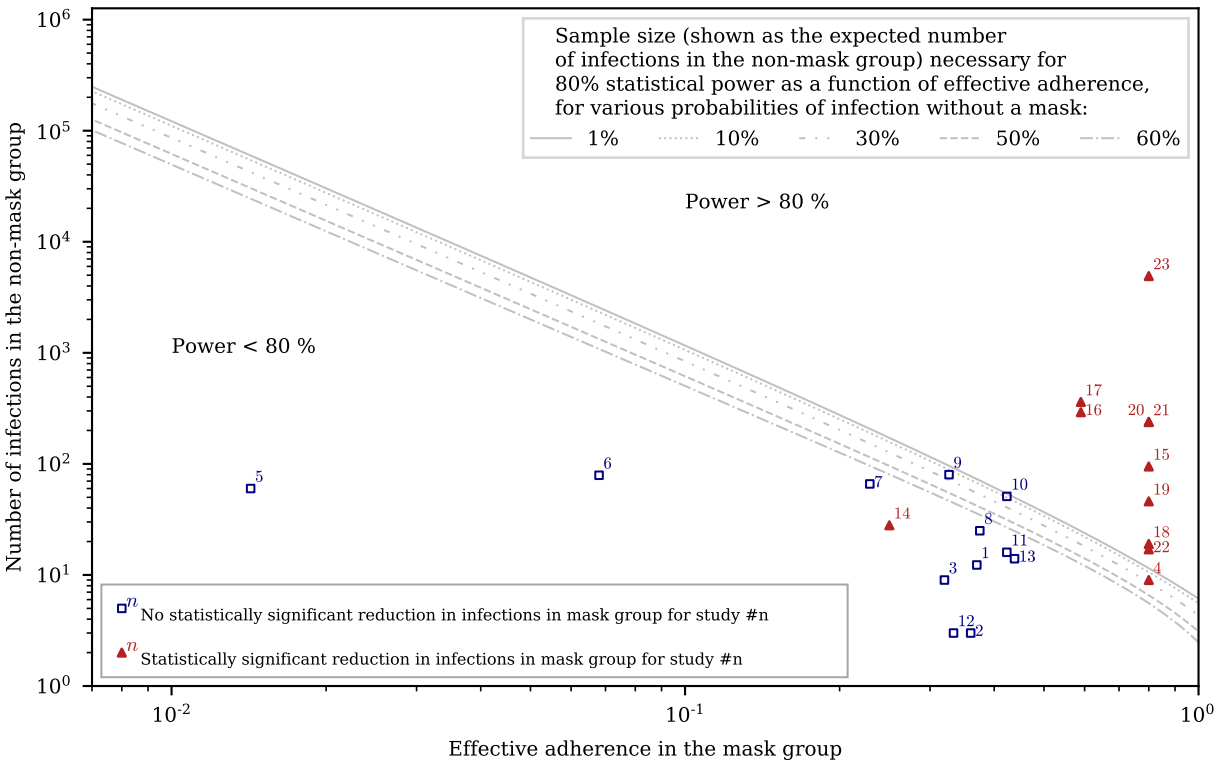


Figure 5-1: The effective adherence and sample sizes of studies that found masks to be effective (red triangles) and those that did not (blue squares). Empirical studies with higher levels of statistical power consistently show that masks protect the wearer; studies with lower statistical power are mixed, as would be expected. The statistical power depends on the sample size, the effective adherence (i.e. mask effectiveness multiplied by the fraction of exposures for which masks are on average worn in the mask group), and probability of being infected without a mask in the setting of the study. Each curve depicts the required sample size (expressed as the expected number of infections in the non-mask group) as a function of effective adherence in order for the study to have a power of 80%, i.e. an 80% probability of finding a statistically significant result ($p < 0.05$, two-tailed). The curves are calculated assuming equal-sized non-mask and mask groups, with each curve representing a different probability of infection in the non-mask group; the total sample size is thus the expected number of infected individuals in the non-mask group (which is what is plotted) divided by the probability of infection and then multiplied by two. We assume that the probability of infection is reduced linearly with increasing effective adherence (see fig. 5-2); this assumption will underestimate the true necessary sample size. The scattered data points depict the size and effective adherence of studies taken from a recent systematic review [69]; the study numbers correspond to those in the first column of fig. 5-5. The effective adherence for the studies are overestimated by assuming that masks are 100% effective; even with this assumption, the studies numbered 1 through 14 were found to have less than 80% statistical power. Note that the power of the studies (reported in fig. 5-5) depend on the probability of infection in the non-mask group and the size of the mask group in addition to the information present in this figure; the location of the studies in this figure relative to the curves should therefore be considered only approximately. In particular, some of the studies have smaller overall sample sizes than implied by this figure, due to their non-mask and mask groups not being equal in size. Mathematical details can be found in the Appendix.

5.2 Statistical power

Some empirical studies find masks to be effective in preventing disease transmission while others do not [159, 232, 103, 8, 360, 69, 93]. However, due to poor statistics, even the studies with negative results are not inconsistent with masks being highly effective. While some of the studies conducted a power analysis to estimate the sample size required to obtain a statistically significant result with 80% probability (i.e. to achieve 80% power, the standard level by convention), these power analyses did not take in to account the possibility of low adherence (i.e. masks being worn for a low percentage of exposure events) and/or the possibility of a very low probability of infection even without a mask. When we consider such factors, none of the studies we analyze that did not find masks to be effective had sufficient statistical power.

The sample of studies we consider is taken from a recent systematic review [69]; see table 5.2 for a list of studies that were excluded and why. Most of the studies we examine measure whether surgical masks protect the wearer; the exceptions are studies nos. 8, 12 and 15, which measure whether masks prevent the wearer from infecting others, and studies nos. 1, 7, 13, and 18, in which both the susceptible and infected individuals sometimes wore masks (see table 5.1).

In order to account for adherence, we make two conservative assumptions that will result in our overestimating the studies' statistical powers. First, we assume that the degree to which a mask reduces the probability of infection is proportional to the fraction of exposures for which it is worn (e.g. we assume wearing a mask half as often provides half as much protection); in fact, wearing a mask half as often will reduce the probability of infection by less than half as much (see fig. 5-2), meaning that we overestimate the statistical power of these studies. Second, the numbers we calculate represent the power the studies would have had were masks 100% effective (i.e. were it impossible to become infected while wearing a mask). To the extent that masks are less than 100% effective, even larger sample sizes would be needed.

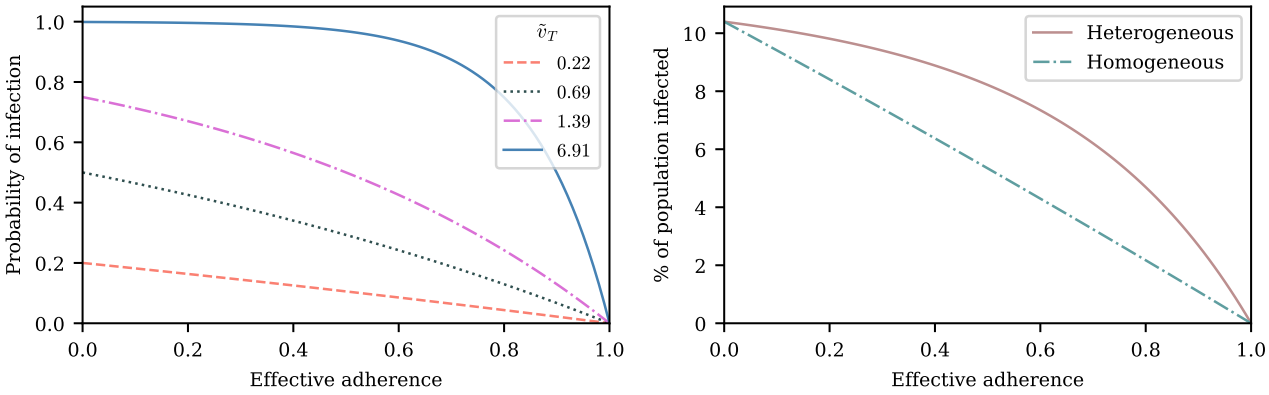


Figure 5-2: **Left:** A susceptible individual's probability of infection as a function of effective adherence $\alpha\gamma$ (mask effectiveness γ multiplied by the fraction of exposures for which the mask is worn α) for various values of that individual's total effective exposure $\tilde{v}_T$ (the total effective exposure is proportional to the number of exposure events). Note that if a mask were 100% effective ($\gamma = 1$), then the effective adherence would simply equal the adherence α. The curves denoting the infection probabilities are given by $1 - e^{-(1-\alpha\gamma)\tilde{v}_T}$ (eq. (5.5)). For high values of $\tilde{v}_T$, the infection probability is nonlinear in the adherence, while for low values of $\tilde{v}_T$, the infection probability decreases approximately linearly with adherence. **Right:** For a group of individuals (e.g. in an arm of a study), the total effective exposure will in general vary from individual to individual such that even if on average the total effective exposure is relatively low, it may be high for the individuals who make up the bulk of those being infected. Thus, while using the average effective exposure would predict an approximately linear decrease of infection probability with increasing adherence, such an approach may overestimate the expected effect of partial mask usage. The dashed curve depicts the expected percentage of infected individuals for the homogeneous case in which everyone experiences the same total effective exposure, whereas the solid curve depicts a case in which the exposure is heterogeneous; in both cases, the percentages of individuals that would be infected without masks (e.g. in a control group) are identical (approximately 10%).

For example, a randomized control trial (RCT) at the Hajj pilgrimage [10] assumed a reduction in infection rate from 12% to 6% in order to determine the sample size necessary for a statistical power of 80%. After taking into account that the randomization was done by cluster (i.e. tent) rather than individual, the required sample size was ~ 6000. However, the study reports that individuals in the intervention group on average wore masks for far less than half the time. Under these conditions, even with perfectly effective masks, a 50% reduction in infection probability is impossible. Adherence can be estimated by the product of the fraction of people who wore masks and the fraction of time during which exposure is possible for which masks were used. The data reported in the study indicate an adherence in the mask group of 3.2%, which could cause at most a 3.2% reduction in the probability of infection. However, the adherence in the control group was 1.8%, meaning that the maximum possible expected reduction in infection between the two groups would be $\frac{0.032-0.018}{1-0.018} = 0.014$ (eq. (5.10)). Thus the effective adherence value used for this study is 0.014. In addition, the probability of infection without masks is reported to be quite low (2%). Under these conditions, the required sample size to achieve the desired statistical power of 80% would be ~ 7.8 million (with individual randomization; with cluster randomization an even larger number of participants is needed). Power analyses for other studies and the methods used are described in the appendix and summarized in figs. 5-1 and 5-5 and table 5.1.

Other factors such as false positives may also limit statistical power. For instance, a recent study [77] conducted in Denmark reported that a mask recommendation did not have a statistically significant effect: in the study's primary composite outcome, 42 vs. 53 people tested positive in the intervention and control groups, respectively. However, the vast majority of these positive results were from antibody tests, and given the antibody test's comparable incidence and false positive rates (approximately 2% and 0.8%, respectively), a substantial fraction of the positive antibody tests in both the control and intervention groups are likely to be false positives, which would affect both the study's power and its statistical analyses [321]. Further false positives could arise from individuals who were infected before the study but for whom seroconversion did not occur until partway through the study. How-

ever, false positives were not accounted for in the study's statistical analysis or conclusions. If only the more reliable PCR tests are considered, then the reduction in infection due to masks (0 vs. 5 infections) is statistically significant ($p < 0.05$).

5.3 Nonlinear effects

We now describe a framework to account for the nonlinear aspects of mask effectiveness. Given that there is a threshold for the viral dose (the amount of the virus inhaled) below which the probability of infection is very small due to the innate immune system [361, 235], and given that the probability of infection p will converge to one (for susceptible individuals) as the viral dose v is increased without limit, the probability of infection as a function of viral dose $p(v)$ is described by a sigmoid function or S-curve (fig. 5-3, see Appendix for details). (Concave curves have also been used to model dose response curves, but such an approach ignores threshold effects [317, 326].) For a single exposure event, we can define the dimensionless *effective exposure* $\tilde{v} \equiv -\ln(1 - p(v))$ such that the probability of infection is $1 - e^{-\tilde{v}}$. Conveniently, the effective exposure is additive for independent exposure events, i.e. the total probability of infection is given by $1 - e^{-\tilde{v}_T}$ where the *total effective exposure* $\tilde{v}_T$ is simply equal to the sum of the effective exposures for each exposure event.

Because the probability of infection is a concave function of the total effective exposure (fig. 5-3), the protection afforded by a mask is super-linear in the percentage of exposures for which it is worn (e.g. wearing a mask twice as often is more than twice as effective; see fig. 5-2). These nonlinear effects can be substantial for high cumulative exposures. Under such conditions, a mask may need to be worn for most or nearly all of the exposure events in order to provide significant protection; otherwise the individual is likely to be infected during the exposures for which the mask is not worn. In the limit of an extremely high total exposure, a mask will of course not have an effect on the probability of infection since a susceptible individual will be infected with nearly 100% probability regardless of whether or not the mask is worn.

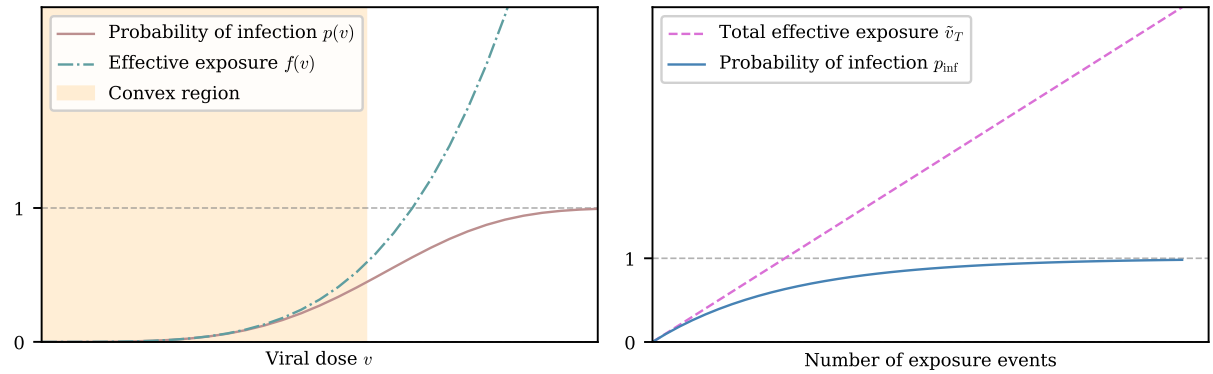


Figure 5-3: **Left:** A representative function for a susceptible individual's probability of infection p as a function of viral dose v for a single exposure event, together with the effective exposure $\tilde{v} \equiv f(v) \equiv -\ln(1-p(v))$. $f(v)$ is convex for all v, while $p(v)$ is convex for sufficiently small v. The convexity of $f(v)$ (which is demonstrated in the Appendix) yields an S-curve for $p(v)$. Note that for any particular viral dose v, the effective exposure $\tilde{v} = f(v)$ can vary from individual to individual. **Right:** A depiction of how the total effective exposure $\tilde{v}_T$ and the probability of eventually becoming infected scale with the number of exposure events. The total effective exposure is the sum of the effective exposures from each exposure event; see Appendix for details.

On the other hand, for low total exposures, the protection masks provide will be approximately proportional to the fraction of exposures for which they are worn. It should be noted that the total exposure of individuals can vary within any given study, such that even if the overall probability of infection is low, most of those who were infected may have been subjected to high cumulative exposures. Studies with low overall probabilities of infection also have an additional difficulty, which is that large sample sizes will be necessary in order that there may be enough infections in the non-mask group to produce a statistically meaningful comparison. In other words, for sufficiently low total exposure, the probability of infection will be quite low even without a mask, and so further reductions to this probability, even if proportionally large, will be small in absolute terms.

We can also analyze certain compound effects that are not considered in most empirical studies. For instance, masks worn on both the infected and susceptible individuals may prevent a transmission event even if neither mask individually would have. Furthermore, this compound effect may be super-linear: if the effect of only an infected individual wearing a mask is to reduce the infection probability by a factor of p_1 and the effect of only a susceptible individual wearing a mask is to reduce the infection probability by a factor of p_2,

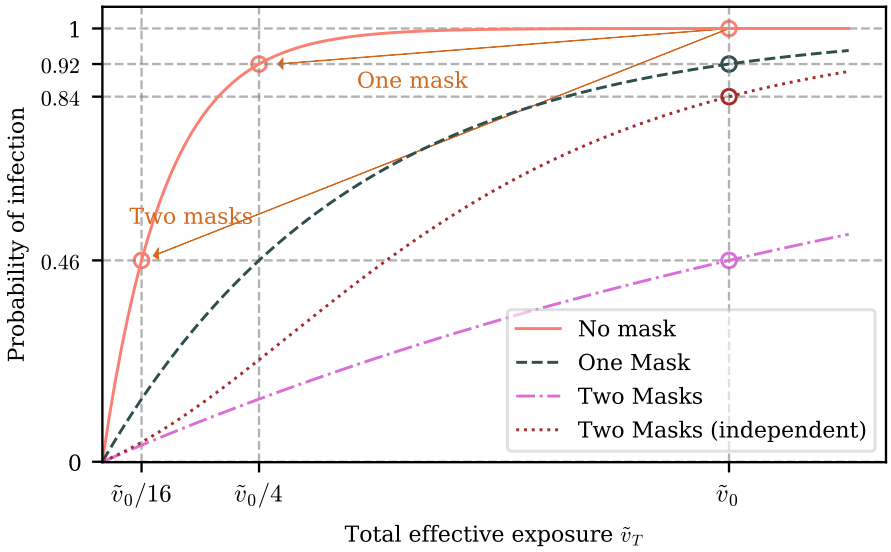


Figure 5-4: The effect of both the susceptible and infected individual wearing a mask can be much larger than the effect of only one of them wearing a mask. In the depicted example, the total effective exposure $\tilde{v}_0$ if neither the infected nor susceptible individual are wearing masks is such that the probability of infection $p_{\text{inf}}(\tilde{v}_0)$ is very close to 1. If each mask reduces the effective exposure by a factor of 4, then the probability of infection if only one of the two individuals is wearing a mask is $p_{\text{inf}}(\tilde{v}_0/4) = 0.92$, i.e. a reduction in risk by a factor of 1.08. If both individuals are wearing a mask, however, the probability of infection is $p_{\text{inf}}(\tilde{v}_0/16) = 0.46$, corresponding to a reduction in risk by a factor of 2.17, which is greater than the product of the effects of each mask individually (shown by the red dotted curve). For illustrative purposes we have assumed that the infectious individual wearing a mask has the same effect as the susceptible individual wearing a mask, but relaxing this assumption will not qualitatively change the results; see Appendix for details.

both individuals wearing a mask could reduce the infection probability by far greater than a factor of $p_1 p_2$, especially for large total effective exposures. In the example shown in fig. 5-4, the probability of transmission is reduced by only a factor of 1.08 (a 7% reduction) due to one or the other individuals wearing a mask, while if both wear a mask, the probability of transmission will be reduced by a factor of 2.17 (a 54% reduction). Similarly, just as there can be a super-linear compound effect from both individuals wearing masks, there can also be super-linear compound effects when mask-wearing is combined with other behaviors that reduce exposure, such as social distancing. Nonlinear effects continue to accumulate when multiple individuals perform multiple behavioral changes that reduce exposure. Recognizing these nonlinear effects is key to appreciating the effectiveness of transmission prevention policies.

It should also be noted that the proportional risk reduction from masks is expected to be large because—due to the convexity of the S-curve (fig. 5-3) when exposure is low (as is likely for many exposure events)—the probability of the mask-wearer being infected is decreased by a greater factor than the decrease in the viral dose [322] (see Appendix). For these low exposure events, although the probability of infection may be small for any given potential transmission event, given multiple events, the large factors by which the probabilities of infection decrease due to this convexity can significantly reduce both the spread of the virus and the probability that the wearer eventually is infected. In other words, wearing a mask may not only prevent the wearer from spreading viruses to others but may also have a surprisingly large protective effect for the mask-wearer. Indeed, studies that analyze population-level data show that masks significantly reduce transmission [178, 230, 196, 212, 88].

In addition to the probability of infection, the implications of a nonlinear dose-response curve apply to several other outcomes as well. In all of the above analyses, the probability of infection can be replaced with the probability of death or the probability of a particular degree of severity of symptoms, each of which can have a unique S-curve (that can also vary from individual to individual). Thus, even when a mask does not prevent infection, it may reduce the severity of symptoms and the chance of long-term health damage or death. It has been observed for the influenza virus that increasing the viral dose may lead to more adverse symptoms [244, 341, 141], an effect that may also apply to SARS-CoV-2 [86, 141, 181, 59].

5.4 Conclusions

Masks block some fraction of viral particles from dispersing from those who are infected and from infecting those who are susceptible and are understood to prevent disease transmission through this mechanism. However, this simple understanding has been questioned based upon mixed empirical evidence. Here we have shown that studies that did not find masks to be effective had too little statistical power to imply anything about the effectiveness of masks.

We have also shown that for many exposure events, masks will reduce the probability of infection by a greater factor than the factor by which they filter viral particles. This effect is also compounded non-linearly when both infected and susceptible individuals wear masks. When interpreted in light of this *a priori* reasoning and the other considerations discussed above, the evidence indicates that, in addition to preventing the wearer from spreading respiratory infections, masks also protect the wearer from contracting them. The studies that did not find statistically significant effects prove only that masks cannot offer protection if they are not worn.

5.5 Appendix

5.5.1 Accounting for non-linearities in the effectiveness of masks

In this section we develop a framework with which to understand the effect of masks. We show that even if masks were to reduce the viral dose by only a modest factor, they may have a significantly larger impact on the probability of infection. We demonstrate that wearing a mask more frequently can super-linearly reduce one's chance of infection (e.g. wearing a mask 80% of the time reduces one's probability of infection by more than twice as much as wearing a mask 40% of the time). We also show that when both infected and susceptible individuals wear masks, there can be a super-linear compound effect (e.g. if only infected individuals wearing masks reduces the probability of infecting susceptible individuals by a factor of 3 and only susceptible individuals wearing masks reduces the probability of being infected by a factor of 2, then if both wear masks the probability of infection will be reduced by a factor that is greater than $2 \times 3 = 6$).

General Framework

Although there is insufficient data to precisely describe the probability of infection as a function of the viral dose inhaled in a single exposure event, we can nonetheless derive some constraints on its shape. For a susceptible individual, the probability of infection (or any

other outcome such as hospitalization or death) p is a function of the viral dose v, i.e. the quantity of virus to which the individual is exposed. (This function $p(v)$ will vary from individual to individual based on biological factors, but should retain the general properties described below.) For small v the probability of a susceptible individual becoming infected will approach zero, and for large v this probability will approach one, so $p(0) = 0$ and $p(\infty) = 1$. Since receiving two viral doses at once should not result in a lower probability of infection than the hypothetical in which the exposure to each viral dose could be modeled as an independent event, we have that

$$p(v_1 + v_2) \geq p(v_1) + p(v_2) - p(v_1)p(v_2) \tag{5.1}$$

Equality will hold only in the absence of threshold effects; given that such effects are well established, we expect the inequality to be strict for small v_1 and v_2. In order to characterize the set of functions satisfying eq. (5.1), we transform $p(v)$ using $p(v) \equiv 1 - e^{-f(v)}$, or equivalently, $f(v) \equiv -\ln(1 - p(v))$. Eq. 5.1 is then equivalent to

$$f(v_1 + v_2) \geq f(v_1) + f(v_2) \tag{5.2}$$

Thus eq. (5.1) is equivalent to $f(v)$ being convex. Choosing a convex $f(v)$ and then transforming back to $p(v)$ yields an S-curve (fig. 5-3), also known as a sigmoid function or sigmoid curve.

When $f(v) \ll 1$, it can be shown by Taylor expansion that $p(v) \approx f(v)$. Thus, for small viral doses, $p(v)$ will be convex as well. If a mask reduces the viral dose v by a factor b [64, 143], then the mask will reduce the probability of infection (or of some other outcome denoted by $p(v)$ such as the probability of severe infection or death) by a factor of $\frac{p(v)}{p(v/b)}$, which depends on v. When $p(v)$ is convex, the factor by which the mask reduces the probability of infection will be greater than b (since convexity implies that $p(v/b) < \frac{1}{b}p(v) + (1 - \frac{1}{b})p(0) = \frac{1}{b}p(v)$). Thus, for small exposures, masks can result in a surprisingly large reduction in

the probability of infection. We treat the impact of masks in more generality below, after introducing a framework for considering multiple exposure events.

The S-curve describes the probability of infection for a single exposure event. For N independent exposure events, the probability of getting infected is $p_{\text{inf}} = 1 - \prod_{i=1}^{N}(1 - p(v_i))$. Using the form $p(v) = 1 - e^{-f(v)}$ as discussed above,

$$p_{\text{inf}} = 1 - e^{-\sum_{i=1}^{N} f(v_i)} \tag{5.3}$$

Defining the *effective exposure* $\tilde{v} \equiv f(v)$ and defining $\tilde{p}(\tilde{v}) \equiv 1 - e^{-\tilde{v}}$,

$$p_{\text{inf}} = \tilde{p}(\tilde{v}_T) \tag{5.4}$$

where $\tilde{v}_T = \sum_i \tilde{v}_i$ is the total effective exposure. Considering the effective exposure $\tilde{v}$ rather than the actual dose v is convenient since the effective exposure for repeated independent exposures is simply the sum of the individual effective exposures. Note that for small effective exposures, the probability of being infected is approximately equal to the effective exposure, i.e. $\tilde{p}(\tilde{v}) \approx \tilde{v}$ for $\tilde{v} << 1$.

One mask

Let γ be the typical amount by which a mask reduces the effective exposure from a single exposure event—i.e. $\tilde{v} \to (1 - \gamma)\tilde{v}$.

Since $f(0) = 0$ and $f(v)$ is convex, the simplest possible expression for $f(v)$ is the scale-free form $f(v) = (v/v_0)^{\beta}$, for some $v_0 > 0$ and $\beta > 1$. In this case, if a mask reduces the viral dose v to v/b, γ can be calculated exactly as $\gamma = 1 - b^{-\beta}$, regardless of v. For small exposures, the infection probability is roughly equal to the effective exposure, which is reduced by a factor greater than b (i.e. $\frac{1}{1-\gamma} > b$) due to convexity ($\beta > 1$), consistent with the analysis above. This effect could potentially be quite large: e.g. for $\beta = 4$, a mask filtering half of the viral particles ($b = 2$) corresponds to a sixteen-fold reduction in effective exposure ($\gamma \approx 0.94$). If we relax this assumption on the function f, then γ becomes an effective parameter that

may depend on the distribution of viral doses to which an individual is exposed. Regardless of the precise form of $f(v)$, however, $\frac{1}{1-\gamma} > b$ will always hold due to the convexity of $f(v)$, i.e. masks will always have a disproportionately large effect on the effective exposure (and thus also on the infection probability when the effective exposure is small).

Then, if a mask is worn for a fraction α of all exposures, the total effective exposure will be reduced from $\tilde{v}_T$ to $(1 - \alpha\gamma)\tilde{v}_T$. The probability of infection is thus

$$\tilde{p}((1 - \alpha\gamma)\tilde{v}_T) = 1 - e^{-(1-\alpha\gamma)\tilde{v}_T} \tag{5.5}$$

(see fig. 5-2).

Thus, we see that for any fixed γ (mask effectiveness) and $\tilde{v}_T$ (total effective exposure without a mask), the benefit of wearing a mask is a convex function of the fraction α of the exposure events for which it is worn. In other words, wearing a mask x times as often will reduce the infection probability by more than a factor of x. Thus, even if masks were 100% effective ($\gamma = 1$), a study in which participants wear masks 10% of the time would need to have sufficient power to detect less than a 10% reduction in the probability of infection. Our analysis therefore overestimates the true power of the studies.

Two masks

To the extent that two masks together have an approximately linear effect on the effective exposure (e.g. if one person wearing a mask reduces effective exposure by $1 - \gamma_1$ and the second person wearing a mask reduces effective exposure by $1 - \gamma_2$, then both wearing masks reduces effective exposure by $1 - \gamma_{12} \approx (1 - \gamma_1)(1 - \gamma_2)$), the effect on the probability of transmission will be super-linear, since the probability of infection $\tilde{p}(\tilde{v})$ is concave in the effective exposure $\tilde{v}$. In other words, especially for individuals who would have received a large total effective exposure without masks, both the susceptible and infectious individuals wearing masks will have a larger effect than would be calculated if each mask had an independent effect on the probability of transmission.

If the effect of the two masks on the effective exposure is super-linear (i.e. $1 - \gamma_{12} < (1 - \gamma_1)(1 - \gamma_2)$), then the effect on the probability of transmission will be super-linear to an even greater extent. If the effect of the two masks on the effective exposure is sub-linear (i.e. $1 - \gamma_{12} > (1 - \gamma_1)(1 - \gamma_2)$), then whether or not they still have a super-linear effect on the probability of transmission will depend on the total effective exposure.

(Note: Under the simplest possible form for $\tilde{v} = f(v)$, i.e. $f(v) = (v/v_0)^\beta$, if the mask on the infected individual reduces v by a factor of b_1, the mask on the susceptible individual reduces v by a factor of b_2, and together the masks reduce v by a factor of $b_1 b_2$, then the masks will have a linear effect on effective exposure, i.e. $1 - \gamma_{12} = (1 - \gamma_1)(1 - \gamma_2)$. Under other forms for $f(v)$ or assumptions about how the masks affect v, other behavior is possible.)

5.5.2 Power analyses

Let p_1 and p_2 be the probabilities of getting infected in the non-mask (size N_1) and mask group (size N_2), respectively. Defining $\epsilon = p_1 - p_2$, the null hypothesis is $H_0 : \epsilon = 0$ and the alternate hypothesis is $H_1 : \epsilon \neq 0$. A test statistic is

$$W = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}_1(1 - \hat{p}_1)/N_1 + \hat{p}_2(1 - \hat{p}_2)/N_2}} = \frac{\hat{p}_1 - \hat{p}_2}{\hat{s}} \tag{5.6}$$

where $\hat{p}_1$ and $\hat{p}_2$ refer to the observed fraction of infections, assumed to be normally distributed random variables whose means are p_1 and p_2 (this approximation is asymptotically exact). We use the shorthand $\hat{s}$ for the denominator of W; note that $\hat{s}$ is an estimator for s, where $s^2 = p_1(1 - p_1)/N_1 + p_2(1 - p_2)/N_2$ is the sum of the asymptotic variances of $\hat{p}_1$ and $\hat{p}_2$. Asymptotically, $W - \epsilon/s$ follows a standard normal distribution. Using the standard notation $\Phi(z_{1-\alpha/2}) = 1 - \alpha/2$ where $\Phi(x)$ is the standard normal cumulative distribution function, the rejection region under H_0 for a significance level of α is given by the union of

$$W < -z_{1-\alpha/2} \text{ and } W > z_{1-\alpha/2} \tag{5.7}$$

No.	Name	Year	Mask Use	Adherence	Size of non-mask group	Fraction of non-mask group infected	Size of face mask group	Fraction of face mask group infected	Statistical Power	Required sample size for a power of 0.8	Actual sample size	Primary Outcome	Significant reduction in infections in mask group?
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)	(XI)	(XII)	(XIII)	(XIV)
1	Cowling (ITTA) [45]	2008		0.37	205	0.06	61	0.07	0.118	2910	266	Antibody Test	No
2	MacIntyre (ITTA) [19]	2009	45% people used masks	0.36	100	0.030	94	0.064	0.078	6444	194	ILI	No
3	MacIntyre (PPA) [19]	2009	Less than 2/5 days	0.32	170	0.053	19	0.211	0.066	4636	189	ILI	No
4	*MacIntyre (PPA) [19]*	2009	5/5 days	0.80	170	0.053	18	0.000	0.298	530	188	ILI	Yes
5	Alfelali (ITTA) [24]	2020	Extremely low	0.014	3139	0.019	3199	0.030	0.051	7862388	6338	PCR	No
6	Alfelali (PPA) [24]	2020	~1.6 hours/day	0.068	2200	0.036	1291	0.051	0.067	176108	3491	PCR	No
7	Simmerman (ITTA) [46]	2011		0.23	292	0.226	291	0.227	0.347	1866	583	ILI	No
8	Canini (ITTA) [48]	2010	3.9 hours/day	0.38	158	0.158	148	0.162	0.345	990	306	Antibody Test	No
9	Aiello (ITTA) [44]	2010	~4 hours/day	0.33	177	0.452	99	0.444	0.699	330	276	ILI (Fever)	No
10	Aiello (ITTA) [43]	2012	~5 hours/day	0.42	370	0.138	392	0.117	0.735	884	762	ILI	No
11	Aiello (ITTA) [43]	2012	~5 hours/day	0.42	370	0.043	392	0.031	0.284	3074	762	PCR	No
12	MacIntyre (ITTA) [47]	2016	~4 hours/day	0.33	295	0.010	302	0.003	0.074	22924	597	ILI	No
13	Cowling (ITTA) [49]	2009		0.44	257	0.054	258	0.070	0.269	2234	515	PCR	No
14	*Barasheed (ITTA) [50]*	2014		0.25	53	0.528	36	0.306	0.236	440	89	ILI	Yes
15	***Sung (pre-post) [51]***	2012	Hospital setting	0.80	920	0.103	454	0.033	1.000	260	1374	PCR	Yes
16	***Choudhry (survey) [52]***	2006	Most of the time	0.59	477	0.612	340	0.150	1.000	52	817	ARI	Yes
17	***Al-Jasser (survey) [53]***	2013	Most of the time	0.59	656	0.550	216	0.454	1.000	62	872	URTI	Yes
18	***Suess (ITTA) [54]***	2012		0.80	82	0.232	69	0.087	0.938	102	151	PCR	Yes
19	***Wu (survey) [55]***	2004	Mask usage outdoors	0.80	73	0.630	70	0.386	1.000	22	143	ILI	Yes
20	***Kim (survey) [56]***	2012	Mask usage in schools	0.80	4164	0.057	466	0.030	1.000	486	4630	ILI	Yes
21	***Lau (survey) [57]***	2004	Mask usage in public	0.80	511	0.466	479	0.192	1.000	38	990	Probable SARS	Yes
22	***Lau (survey) [58]***	2004	Mask usage in hospitals	0.80	98	0.173	177	0.079	0.927	144	275	Probable SARS	Yes
23	***Wu (survey) [59]***	2016	Mask usage in hospitals	0.80	10298	0.477	2728	0.423	1.000	36	13026	ILI	Yes

Figure 5-5: Summary of statistical power analysis. Given the adherence levels reported in the studies the sample size necessary for a statistical power of 80% for a two-tailed test and significance level of 0.05 (assuming participants are equally divided between the non-mask and mask groups) is presented in column XI. The statistical power given the actual sizes of the non-mask and mask groups is presented in column X. These calculations were made for the case in which masks are 100% effective; if masks are effective but not perfectly so, the necessary sample sizes for 80% power (column XI) will be larger, while the statistical powers given the actual sample sizes (column X) will be lower. Studies found to have greater than 80% power are in bold (nos. 15-23), and studies that found a statistically significant reduction in infections in the mask group are italicized (nos. 4, 14-23). Adherence is defined as the fraction of exposure for which masks were used; calculations of adherence for each study are presented in table 5.1. For studies that reported multiple analyses, each analysis is listed as its own entry (e.g. Aiello (2012) [5] performed one analysis in which infection is defined by influenza-like illness (no. 10) and one analysis in which infection is defined by a positive PCR test result (no. 11)). Note that only in the intention-to-treat analyses are participants randomly divided between the non-mask and mask groups; in survey and per-protocol analyses, which group a participant belongs to depends on whether or not that individual reported wearing a mask with a frequency above a threshold decided by the study.
Abbreviations: ITTA: Intention-to-treat analysis; **PPA**: Per-protocol analysis; **ILI**: Influenza-like illness; **ARI**: Acute respiratory infection; **URTI**: Upper respiratory tract infection; **PCR**: Polymerase chain reaction test (nasopharyngeal swab test); **SARS**: Severe Acute Respiratory Syndrome

The various studies may use slightly different statistical tests, but the differences between tests should be small and will asymptotically disappear entirely. For any particular values of ϵ and s, the probability $W < -z_{1-\alpha/2}$ is asymptotically given by $\Phi(-z_{1-\alpha/2} - \epsilon/s)$ and the probability $W > z_{1-\alpha/2}$ is asymptotically given by $1 - \Phi(z_{1-\alpha/2} - \epsilon/s) = \Phi(-z_{1-\alpha/2} + \epsilon/s)$. Thus, given ϵ and s, the power, denoted by $1 - \beta$ and equal to the probability that the null hypothesis is rejected if it is indeed false, is asymptotically given by

$$1 - \beta = \Phi(-z_{1-\alpha/2} - \epsilon/s) + \Phi(-z_{1-\alpha/2} + \epsilon/s) \tag{5.8}$$

Under the assumptions that masks are fully effective ($\gamma = 1$) and that the probability of infection p_{inf} decreases linearly with the adherence, the effect of mask usage is

$$p_{\text{inf}} \to p_{\text{inf}}(1 - a) \tag{5.9}$$

where the adherence a is the average fraction of exposure events for which the masks were used (see section 1 of the Appendix; here we use a instead of α for the adherence to avoid confusion with the significance level). Thus, for an infection probability $p_{\text{inf}} = p_1$ in the non-mask group (size N_1), the infection probability in the mask group (size N_2) will be $p_2 = p_1(1 - a)$. Thus, by estimating p_1 and a for each study, we can use eq. (5.8) to find power of each study given the sizes of their non-mask and mask groups, as well as the sample size (i.e. total number of participants) that would have been required for 80% power. For the latter estimate, we assume a study design in which the participants are evenly divided between the non-mask and mask groups (i.e. $N_1 + N_2 = 2N_1 = 2N_2$) and rounded up the necessary sample size to the nearest even integer.

For certain studies, some participants in the non-mask group used masks as well. In this case, adherence in both the mask group and non-mask group must be considered. Under the assumption that probability of infection decreases linearly with effective adherence, the probability of infection in the non-mask group p_1 is related to the probability of infection without masks p_0 by $p_1 = p_0(1 - \gamma a_1)$ where a_1 is the adherence in the non-mask group

and γ is mask effectiveness. Then the probability of infection in the mask group will be $p_2 = p_0(1 - \gamma a_2)$ where a_2 is the adherence in the mask group. The net adherence a is defined by $p_2 = (1 - \gamma a)p_1$, which yields $a = \frac{a_2 - a_1}{1 - \gamma a_1}$. In our analyses we assume $\gamma = 1$, which leads to an overestimate for the net adherence a of

$$a = \frac{a_2 - a_1}{1 - a_1} \tag{5.10}$$

We estimate p_1 using the observed fraction of infections in the non-mask group $\hat{p}_1$. To check the robustness of our conclusions, we did a sensitivity analysis and found that if $\hat{p}_1$ differs from p_1 by a standard deviation (i.e. if we increase our estimate of p_1 by $\sqrt{\frac{1}{4N_1}}$), all studies that were under-powered ($< 80\%$), except for one [5] remain under-powered. (To ensure robustness we used $\sqrt{\frac{1}{4N_1}}$ as the standard deviation, which is the maximum possible value of the true standard deviation $\sqrt{p_1(1 - p_1)/N_1}$.) If $\hat{p}_1$ underestimates p_1 by two standard deviations, another study [4] would have greater than 80% power under our assumptions. It should be noted, however, that these assumptions overestimate the power in multiple ways (fully effective masks, overestimated adherence values, assuming a linear relationship between adherence and effectiveness, and the fact that individuals whose infections were not detected until after the start of the study could have actually been infected before they start of the study, i.e. before the mask intervention was implemented).

A more significant limitation of our analysis is in the difficulty in estimating adherence. Adherence is often reported qualitatively, and even when quantitative, it is reported as the amount of time for which one wears a mask, which may differ from the fraction of exposures for which masks were worn. To account for this difficulty, our strategy has been to consistently overestimate statistical power; to this end, we have erred on the side of overestimating adherence (see table 5.1), and have also used other overestimating assumptions described in the previous paragraph.

Table 5.1: Adherence calculations for each study.

No.	Name	Year	Masks used by	Description and calculation
1	Cowling (ITTA) [105]	2008	Infected patients and their contacts	Household study: 45% of 21 index cases used masks and 21% of 61 contacts wore masks. To overestimate adherence, we assume no transmission occurs while either the index patient or contact is wearing a mask. Neglecting correlations between whether or not the index patient wore a mask and the number of contacts of that index patient, an upper bound for the probability that either a contact or the index patient corresponding to that contact used a mask is 45%+21% = 66% (this is likely an overestimate since households in which index patients wear masks and households in which contacts wear masks are almost certainly not mutually exclusive). In the control group, 30 % of index patients and 1 % of contacts used masks. Those classified as using masks used them often or always; therefore we assume that they used masks for 80% of all exposures, a likely overestimate since the participants were asked to use masks only when they are not sleeping or eating. Therefore, the adherence in the mask group is estimated as $0.66 \times 0.8 = 0.53$, and adherence in the control group is estimated as $0.31 \times 0.8 = 0.25$. This leads to a net adherence of 0.37 according to eq. (5.10).
2	MacIntyre (ITTA) [232]	2009	Contacts of infected patients	Household study over 5 days: Contacts were told to use masks when in the same room as the index patient. We consider only the surgical mask group (the other group was using P2 masks). On day 3, maximum adherence was reported: 45% of contacts used masks for most of the time. We assume that those who used masks used them for 80% of exposures, a likely overestimate since contacts did not use masks while sleeping, even if the child (index patient) was next to them in bed, and because the contacts could have been infected even if they were not in the same room as the index patient. The adherence is estimated as $0.45 \times 0.8 = 0.36$
3	MacIntyre (PPA) [232]	2009	Contacts of infected patients	Household study over 5 days (see row no. 2): Participants in this arm of the per-protocol analysis used masks for < 2 out of 5 days. Overestimating adherence at 0.8 for 2 days gives adherence $= 2/5 \times 0.8 = 0.32$.
4	MacIntyre (PPA) [232]	2009	Contacts of infected patients	Household study over 5 days (see row no. 2): Participants in this arm of the per-protocol analysis used masks for all 5 days. Overestimating adherence at 0.8 for 5 days gives adherence an adherence of $5/5 \times 0.8 = 0.8$.

5	Alfelali (ITTA) [10]	2020	Susceptible individuals	Hajj study: From figure 2 of the study, we can only obtain approximate numbers since numerical data is not available in the figure. An average across the four days gives us the percentage of people using masks for various amounts of time. Using the upper bounds of the reported time ranges, we compute the average mask usage duration. For the last time range (greater than 3 hours), we assume that masks were used on average for 5 hours. This leads to an average mask use of 0.778 hours and an adherence in the mask group of $0.778/24 = 0.032$. Participants in the control group used masks for 0.438 hours on average, yielding an adherence in the control group of $0.438/24 = 0.018$. The net adherence value is thus 0.014 (eq. (5.10)). Note that the systematic review [69] uses an older pre-print version of this study.
6	Alfelali (PPA) [10]	2020	Susceptible individuals	Hajj study (see row no. 5): Those who wore masks were compared to those who did not. The average mask use among those who wore masks was 1.637 hours; thus adherence = $1.637/24 = 0.0682$.
7	Simmerman (ITTA) [301]	2011	Infected patients and their contacts	Household study: We compare the hand-hygiene group with the hand-hygiene + mask group. Only median (and not mean) mask usage was reported for the index and contact individuals; we therefore approximate the mean with the median. The median mask usage for the index patient was 35 minutes. The mean of median mask usage for contacts—parents, siblings and other relations—was 107.9 minutes. We estimate that index patients and contacts were in contact for 10.4 hours per day using data from a similar study [233] (row no. 12). Adherence is therefore estimated as $\frac{107.9+35}{60\times 10.4} = 0.23$ (see row no. 1 for why the index and contact mask usages were added together), a likely overestimate, given that the majority of the households resided in small one-bedroom apartments and thus were likely in contact for significantly greater than 10.4 hours per day on average. Furthermore, contacts could have been infected outside of their homes. Also, it was reported that 17.6% of individuals in the control group used masks, meaning it was likely that those in the hand-hygiene-only group did as well (which would further reduce the net adherence).
8	Canini (ITTA) [81]	2010	Infected patients	Household study: Average mask use was 3.9 hours per day (from table 3 of the study). We estimate that index and contact patients were in contact for 10.4 hours per day using data from a similar study [233] (row no. 12). Adherence is therefore estimated as $\frac{3.9}{10.4} = 0.38$, a likely overestimate given that contacts could have been infected outside their homes, or in their homes while not in contact with the index patient.
9	Aiello (ITTA) [4]	2010	Susceptible individuals	University residence hall: Mask usage was recorded inside the residence hall and they were used for 3.92 hours per day. Assuming that residents spent 12 hours outside the halls, we exclude it from the adherence calculation. Adherence = $\frac{3.92}{24-12} = 0.33$, a likely overestimate because participants were only encouraged but not required to use masks outside the residence halls, where they may be infected. In addition, the participants had left the residence halls for spring break, during which they were not required to wear masks.

10	Aiello (ITTA) [5]	2012	Susceptible individuals	University residence hall: Masks were used for 5.08 hours per day. Adherence = $\frac{5.08}{24-12} = 0.42$ (see row no. 9).
11	Aiello (ITTA) [5]	2012	Susceptible individuals	University residence hall: Masks were used for 5.08 hours per day. Adherence = $\frac{5.08}{24-12} = 0.42$ (see row no. 9).
12	MacIntyre (ITTA) [233]	2016	Infected patients	Household study: In the mask group, index patients were in contact with contacts for an average of 10.4 hours, and used masks for an average of 4.4 hours. The adherence in the mask group is thus estimated as $\frac{4.4}{10.4} = 0.42$. In the control group, average mask usage was 1.4 hours; adherence in the control group is thus estimated as $\frac{1.4}{10.4} = 0.13$. Net adherence is thus 0.33 (eq. (5.10)).
13	Cowling (ITTA) [104]	2009	Infected patients and their contacts	Household study: We compare the hand-hygiene group with the hand-hygiene + mask group. In the hand-hygiene + mask group, 49% of index cases and 26 % of contacts used a mask often or always. We therefore calculate adherence in the hand-hygiene + mask group as $(0.49 + 0.26) \times 0.8 = 0.60$ (see row no. 1). In the hand-hygiene group, 5 % of contacts and 31 % of index cases used masks, which leads to an adherence $= (0.31 + 0.05) \times 0.8 = 0.29$ in the hand-hygiene group. Net adherence is thus 0.44 (eq. (5.10)).
14	Barasheed (ITTA) [53]	2014	Susceptible individuals	Hajj pilgramage: 36 people were in the face mask group: 8 people never used a mask; 11 people used masks for < 4 hours; 8 people used masks used for 5-8 hours; 9 people used masks for > 8 hours (from table 2 of the study). Using the upper limits of the duration ranges (and 12 hours for the > 8 hour group), adherence $= \frac{1}{36}(8 \times 0/24 + 11 \times 4/24 + 8 \times 8/24 + 9 \times 12/24) = 0.25$.
15	Sung (Pre-post) [314]	2012	Potentially infected individuals	Visitors had to use face masks when they visited patients in their rooms and the incidence of infections was recorded among the patients. Although adherence was not reported, it is reasonable to assume that adherence was high since the study was conducted in a hospital where doctors and health care workers would have ensured that protocols are followed; in addition the visitors were in contact with patients only for a limited duration. We therefore assume an adherence of 0.8
16	Choudhry (Survey) [91]	2006	Susceptible individuals	Survey study for Hajj pilgrims: We consider the group of male pilgrims who reported using masks most of the time, compared to a group who did not use masks. We assume that masks were not used while sleeping or eating, and note that the pilgrims remain susceptible to infection during such activities since they slept in shared tents. Allotting 10 hours per day for sleeping and eating and other activities during which masks were not worn, we estimate the adherence as $14/24 = 0.59$.
17	Al-Jasser (Survey) [7]	2013	Susceptible individuals	Survey study for Hajj pilgrims: We consider the group of male pilgrims who reported using masks most of the time, compared to a group who did not use masks, and therefore estimate adherence as 0.59 (see row no. 16).

18	Suess (ITTA) [313]	2012	Infected patients and their contacts	Household study: From figures 2 and 3 in the study, the average mask usage (across 8 days and across both seasons) among contacts and index patients is 69.4% and 56.4%, respectively. From this data it is not impossible that in every household either the index patient or contacts were wearing masks; using this potential overestimate, we calculate adherence as $1 \times 0.8 = 0.8$ (see row no. 1).
19	Wu (survey) [356]	2004	Susceptible individuals	Survey study: Face mask usage was reported only outside the home. Adherence was reported subjectively – 'Never', 'Sometimes', 'Always' (table 1 of the study). We compare the groups which used masks always and never used masks, and use an adherence value of 0.8 for the 'Always' group, a likely overestimate since participants could have been infected from household contacts.
20	Kim (survey)[197]	2011	Susceptible individuals	Survey study among school children for influenza: Mask usage during school hours was reported as 'continuous', 'irregular', 'not used'. We assume an adherence of 0.8 for the 'continuous' group (and compare the infection rate to the group that did not use masks), a likely overestimate since children could be infected outside of school hours.
21	Lau (survey) [210]	2004	Susceptible individuals	Survey study during SARS epidemic: Mask usage was recorded only for public places. The study considered the frequent use of masks as using a mask, and occasional/seldom/no use was considered as not using a mask. We assume a value of 0.8 for adherence, a likely overestimate since people could have gotten infected at home where mask usage was not recorded and since some mask usage was possible in the non-mask group.
22	Lau (survey) [209]	2004	Susceptible individuals	Survey study during SARS epidemic: Mask usage was recorded only during hospital visits to patients with SARS. We use an adherence value of 0.8 for hospital settings (see row no. 15). For this study 0.8 is likely an overestimate since SARS infection could have occurred outside of the hospital as well.
23	Wu (survey) [358]	2016	Susceptible individuals	Survey study for influenza-like illness. Mask usage was recorded only during hospital visits. We use an adherence value of 0.8 for hospital settings (see row no. 15). For this study 0.8 is likely a substantial overestimate since infection could have occurred outside of the hospital as well.

Table 5.2: Studies not included in power analysis.

Name	Year	Reason for exclusion from power analysis
Shin [296]	2018	Study was randomized for testing a common cold drug rather than mask usage, and mask usage was comparable in both of the groups.
Zhang [363]	2013	Unknown adherence and incomplete data.
Jolie [190]	1998	Animal to human transmission: We consider only human to human transmission for our analysis.
Tahir [318]	2019	Animal to human transmission: We consider only human to human transmission for our analysis.
Larson [208]	2010	Mask adherence was reported to be 'poor' but neither the percentage of participants using masks nor the duration of mask usage was reported, so we could not make an estimate for the adherence.
Emamian [129]	2013	Survey study for Hajj pilgrims: Adherence for mask usage was reported only as 'Yes' or 'No'. Even occasional use of mask was considered as 'Yes'. Since adherence data stratified by frequency and/or duration was not reported, we could not make an estimate for the adherence.
Deris [117]	2010	Survey study for Hajj pilgrims: Adherence for mask usage was reported only as 'Yes' or 'No'. Since adherence data stratified by frequency and/or duration was not reported, we could not make an estimate for the adherence.
Uchida [332]	2017	Survey study for children. Mask usage was reported as 'using masks at any time or place'. Since adherence data stratified by frequency and/or duration was not reported, we could not make an estimate for the adherence.
Balaban [29]	2012	Survey study for Hajj pilgrims: Adherence for mask usage was reported only as 'Yes' or 'No'. Since adherence data stratified by frequency and/or duration was not reported, we could not make an estimate for the adherence.
Zein	2002	Study not available.

Chapter 6

Negative Representation and Instability in Democratic Elections

We now shift our focus to political systems. This chapter considers the properties of democratic elections; the next chapter will consider consequences of geographic patterns in political opinion. As in previous chapters, the focus has been on identifying the largest-scale behaviors—in this case, by abstracting away the details of the electoral process and considering an election as a map from the political preferences of citizens to those of the elected official. Details are then later considered in light of these considerations. The behaviors identified—instability and negative representation—had not been previously considered in the formal literature due to a focus on mathematical tractability and model building (via assumptions of concave voter preferences) rather than on the various regimes that could arise in the space of possible election behaviors.[1]

[1]What follows is a slightly modified version of the preprint `https://arxiv.org/abs/1810.11489`, which was subsequently published as Alexander F. Siegenfeld and Yaneer Bar-Yam. Negative representation and instability in democratic elections. *Nature Physics* **16**, 186-190 (2020).

6.1 Introduction

The challenge of understanding the collective behaviors of social systems can benefit from methods and concepts from physics [1, 285, 308, 84, 199, 274], not because humans are similar to electrons, but because certain large-scale behaviors can be understood without an understanding of the small-scale details [50], in much the same way that sound waves can be understood without an understanding of atoms. Democratic elections are one such behavior. Over the past few decades, physicists have explored scaling patterns in voting and the dynamics of political opinion formation, e.g. [139, 140, 65, 87, 138, 68]. Here, we define the concepts of *negative representation*, in which a shift in electorate opinions produces a shift in the election outcome in the opposite direction, and *electoral instability*, in which an arbitrarily small change in electorate opinions can dramatically swing the election outcome, and prove that unstable elections necessarily contain negatively represented opinions. Furthermore, in the presence of low voter turnout, increasing polarization of the electorate can drive elections through a transition from a stable to an unstable regime, analogous to the phase transition by which some materials become ferromagnetic below their critical temperatures. Empirical data suggest that United States presidential elections underwent such a phase transition in the 1970s and have since become increasingly unstable.

Elections are, fundamentally, a means of aggregating many opinions into one—those of the citizens into that of the elected official. Here, an opinion refers to a person's entire set of political beliefs; i.e. each citizen (and candidate) has one opinion. For simplicity, we focus on the case in which the set of all possible opinions can be embedded in a one-dimensional continuous space (e.g. a position on a left-right spectrum), as in many studies in the social choice literature (reviewed in section 6.7.3). However, our results can be extended to a multidimensional space (see section 6.7.1). The first part of this chapter examines general properties of electoral representation (which we connect to the voting power literature in section 6.7.2) and instability, using a mathematical formalism that departs from the literature in that it makes no assumptions about how people vote or even the structure of the voting process. The second part of this chapter presents a specific model that builds upon the

social choice literature in order to demonstrate how, when the assumption of concave voter preferences is relaxed, instability and negative representation can arise. The specific model is shown to map onto the well-known mean-field Ising model [192] of magnetic materials. The emergence of instability can couple to geospatially varying local election outcomes and the potentially destabilizing effects of a two party system. Finally, this chapter considers the implications of such a model for contemporary American elections.

6.2 General properties of elections: representation and (in)stability

We define an *election* by $y[f(x)]$, a *functional* that maps the distribution of electorate opinions $f(x)$—defined so that for any interval $[a, b] \subset \mathbb{R}$, the number of citizens with opinions in that interval is $\int_a^b f(x)dx$—to the election outcome $y \in \mathbb{R}$, i.e. the opinion of the elected official. Note that in this framework, candidacy is endogenous: candidate opinions—or equivalently, which candidates run—are themselves functions of the electorate opinions. Any electoral system, regardless of its detailed mechanisms (e.g. the number of candidates or parties, the existence of primary elections, restrictions on candidate entry, etc.) can be conceptualized as such a process that outputs the opinion of the winning candidate based on the electorate opinions. In order that no opinion be *a priori* privileged over others by the voting system itself, the one restriction we place on this process is *translational invariance*, i.e.

$$y[f(x + c)] + c = y[f(x)] \tag{6.1}$$

for all c (section 6.6.1).

In order to measure the sensitivity of the election outcome to changes in electorate opinions [253], we define the *representation* of an opinion x by

$$r_c(f, x) = \frac{\delta y}{c} \tag{6.2}$$

where δy is the change in outcome that occurs if an individual's opinion changes from x to $x' = x+c$ (Which could affect not only which candidate wins the general election, but also the precise positions of the candidates who decide/are selected to run). (Representation should *not* be defined using only the distance between a citizen's opinion and that of the elected candidate: opinions without causal influence on the election outcome are not represented, even if they happen to align with that outcome.) It is convenient to measure representation by

$$r(f, x) = \lim_{c \to 0} r_c(f, x) \tag{6.3}$$

when the limit exists, since $r(f, x)$ does not depend on c.

For a large population, a number of results hold if the election is differentiable (see section 6.6.2 for details). First,

$$r(f, x) = \frac{d}{dx} \frac{\delta y}{\delta f(x)} \tag{6.4}$$

Second, $r_c(f, x)$ is the average of $r(f, x)$ over the interval $[x, x + c]$, and thus representation of individual opinions can be measured by $r(f, x)$ alone. Third, the total representation of the electorate's opinions equals 1:

$$\int_{-\infty}^{\infty} f(x) r(f, x) dx = 1 \tag{6.5}$$

We now show that all unstable elections contain negatively represented opinions. An election is *unstable* if an arbitrarily small change in opinion can cause a sizable change in the election outcome (fig. 6-1), i.e. for some f and x,

$$\lim_{c \to 0} c r_c(f, x) \neq 0 \tag{6.6}$$

If an election $y[f]$ is unstable for an opinion distribution f_0, then if some opinion x_0 changes by a small amount ϵ (call the resulting opinion distribution f_1), the election outcome changes by a larger amount C, i.e. $\delta y_1 \equiv y[f_1] - y[f_0] = C$ with $|C| > |\epsilon|$. Now consider starting with f_0 and shifting all opinions except x_0 by $-\epsilon$ (call the resulting opinion distribution f_2). Since

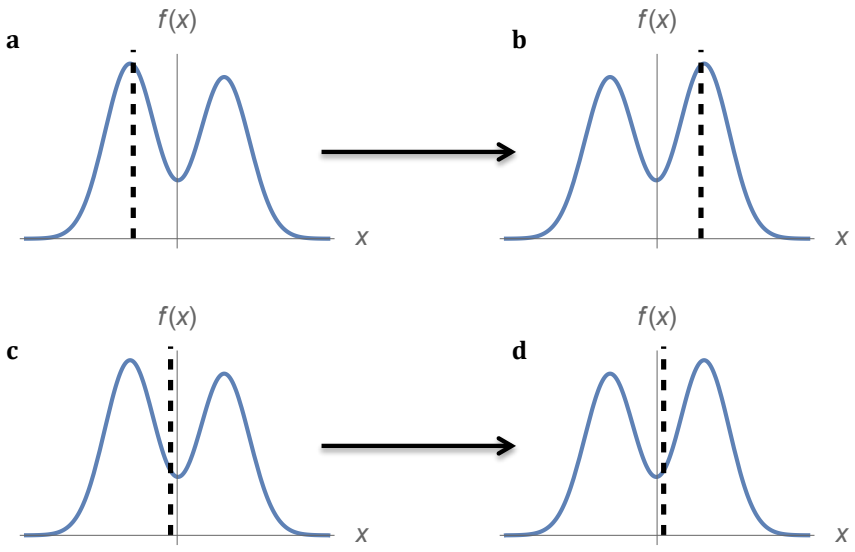


Figure 6-1: Election outcomes (vertical dashed lines) shift in response to a shift in the distribution of electorate opinions ($f(x)$, denoted by the solid curves, where the horizontal axes (x) denote political opinion). **a-b**, When not everyone votes, an election can be unstable (eq. (6.11))—a small shift in opinions to the right causes a large swing in the election outcome. **c-d**, In a stable election, by contrast, a small shift in opinions causes a similarly small shift in outcome.

$f_2(x) = f_1(x + \epsilon)$, $\delta y_2 \equiv y[f_2] - y[f_0] = C - \epsilon$ by translational invariance (eq. (6.1)). Thus, δy_1 and δy_2 have the same sign (since $|C| > |\epsilon|$), despite being caused by changes of opinion in opposite directions, so one of the two changes in opinions must be negatively represented. This proof that unstable elections always contain negatively represented opinions relies only on translational invariance; therefore, we expect it to hold generally in real-world conditions, regardless of the size of the electorate, the number of candidates or parties, the existence of primaries, the effects of the Electoral College, etc.

6.3 Applications

Thus far we have presented general results that apply to all elections; we now examine a particular class of models in order to illustrate how negative representation can arise and to show how polarization drives elections through a phase transition into the unstable regime. Consider an election with two candidates (or two major political parties) who choose their positions in order to maximize their chances of winning the election (or, equivalently, choose to run/are selected to run in the general election based on these considerations). Then, under the assumptions of an affine linear *utility difference model* [39] (see section 6.6.4), the possible election outcomes are given by (Theorem 4 of [39]):

$$\underset{y \in \mathbb{R}}{\operatorname{argmax}} \int_{-\infty}^{\infty} u_x(y) f(x) dx \tag{6.7}$$

where the utility function u_x denotes the political preferences of someone with opinion x. We note that this model always admits at least one Nash equilibrium in candidate strategies (section 6.6.3); the instability we describe, which arises from multiple Nash equilibria, should be distinguished from scenarios in which no Nash equilibria exist (section 6.7.3). Although this model may not accurately describe individual and candidate behaviors, it can nonetheless be used to describe the properties of real-world elections, since representation and instability depend on electoral mechanisms only through the effects of these mechanisms on $y[f]$—the functional that characterizes how the election outcome varies with electorate opinions. For

instance, this model cannot describe the deterministic voting underlying the Median Voter Theorem [63] but can nonetheless exactly capture the collective behavior to which such voting gives rise, namely the election of the median opinion (section 6.6.4).

6.3.1 Low voter turnout

If citizens with opinions that are far from both candidates are more likely to abstain from voting, a phenomenon known as *alienation* [175, 173, 306, 268, 2], then negative representation occurs. Since not voting for either candidate is equivalent to preferring them equally, a utility function $u_x(y)$ that captures this behavior will be almost flat for large $|y - x|$. One such utility function is

$$u_x(y) = u(y - x) = e^{-\frac{(y-x)^2}{2a^2}} \tag{6.8}$$

where a is a positive constant. Representation is then given by (section 6.6.5)

$$r(f, x) \propto -\frac{1}{N}u''(y^* - x) = \frac{1}{Na^4}(a^2 - (y^* - x)^2)e^{-\frac{(y^*-x)^2}{2a^2}} \tag{6.9}$$

where N is the number of constituents. Opinions far from the election outcome ($|x - y^*| > a$) are negatively represented ($r(f, x) < 0$, see fig. 6-2): the election outcome is inversely sensitive to changes in those opinions. For instance, given a center-left candidate and a candidate to the right, as left-wing individuals move farther left, they may become less likely to vote for the center-left candidate (choosing instead not to vote), which increases the probability that the candidate on the right will win. In response, the electoral equilibrium of future elections may shift rightward, as candidates no longer vie for these left-wing votes. (Eq. 6.8 is just an example; more generally, negative representation will occur if and only if $u(y - x)$ is not concave.) Thus, individual choices to abstain when neither candidate is appealing lead to a system-level perversion in the aggregation of electorate opinions, in which the electorate becoming more left-wing can result in a more right-wing outcome, or vice versa.

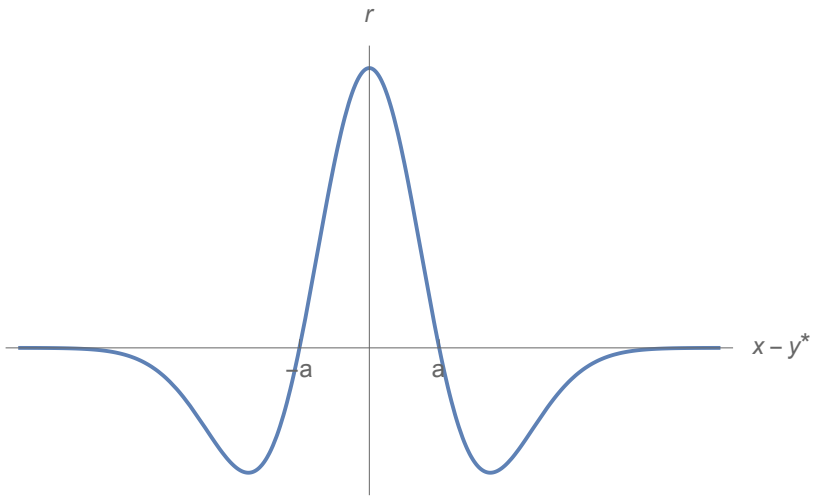


Figure 6-2: When some voters abstain, opinions far from the election outcome may be negatively represented. This graph depicts the representation (r) of opinions (x) as a function of their distance from the election outcome (y^*) for voting behavior given by eq. (6.8).

6.3.2 A phase transition to instability

When combined with a sufficiently polarized electorate, nonvoting (in particular, non-concave u) leads not only to negative representation but also to instability. (As proven earlier, instability cannot occur without negative representation; however, negative representation is possible without instability—see section 6.6.6 for further discussion.) For u_x defined by eq. (6.8) and an opinion distribution $f(x)$ consisting of two normally distributed (potentially unequally sized) subpopulations centered at $\pm\Delta$ (without loss of generality, we define the origin as the (unweighted) average of the means of the subpopulations),

$$f(x) = w_1 e^{-\frac{(x+\Delta)^2}{2\sigma^2}} + w_2 e^{-\frac{(x-\Delta)^2}{2\sigma^2}}, \tag{6.10}$$

the outcome y is given by the following condition:

$$y/\Delta = \tanh(Jy/\Delta + h) \tag{6.11}$$

where J, a dimensionless measure of the polarization of the electorate, is given by $J = \Delta^2/(a^2 + \sigma^2)$ and h, a measure of the relative sizes of the two subpopulations of the elec-

torate, is given by $h = \frac{1}{2} \ln \frac{w_2}{w_1}$. For $w_1 = w_2$ (i.e. for equally sized subpopulations), $h = 0$, and the election is stable for $J \leq 1$ and unstable for $J > 1$. In the stable regime, $y = 0$. In the unstable regime, there are two possible outcomes described by $\pm|y^*|$, and an arbitrarily small change in h can cause y to swing between its positive and negative values (fig. 6-3). Thus, small variations determine which subpopulation wins. The variations might include changes in population, nuances in the candidates' personalities, changes in the rules (simple majority versus Electoral College system, for example), voting restrictions, and the effectiveness of turnout operations. From election cycle to election cycle, the outcome can swing between the two subpopulations, with the majority of the opinions in the losing subpopulation being negatively represented (section 6.6.6). An intuitive explanation of how such negative representation and instability arises is that, due to low voter turnout, candidates are incentivized to focus on turning out their bases rather than winning over centrist voters.

This voting model (eq. (6.11)) is precisely equivalent to a mean-field Ising model of a ferromagnet [192], in which each spin (magnetic dipole) interacts with every other spin, highlighting the hidden dependencies that can arise from a system in which citizens' behaviors are superficially independent (section 6.6.7). This connection with the Ising model differs from other applications of the Ising model to complex systems [40, 247, 158, 304] in two respects: first, our analysis takes the electorate opinions as given and therefore concerns the instability in the election dynamics rather than in the dynamics of the electorate opinions themselves, and second, we do not impose any of the assumptions of the Ising model, but rather show (section 6.6.7) that such an equivalence naturally arises for certain classes of voting behavior. Because the citizens are effectively coupled through collectively choosing a candidate, the system exhibits emergent behavior in which there can be discontinuities in the election outcome despite the continuous voting behaviors of individuals. Such emergent discontinuities are a common feature of complex systems [40, 239, 288, 252, 67, 164].

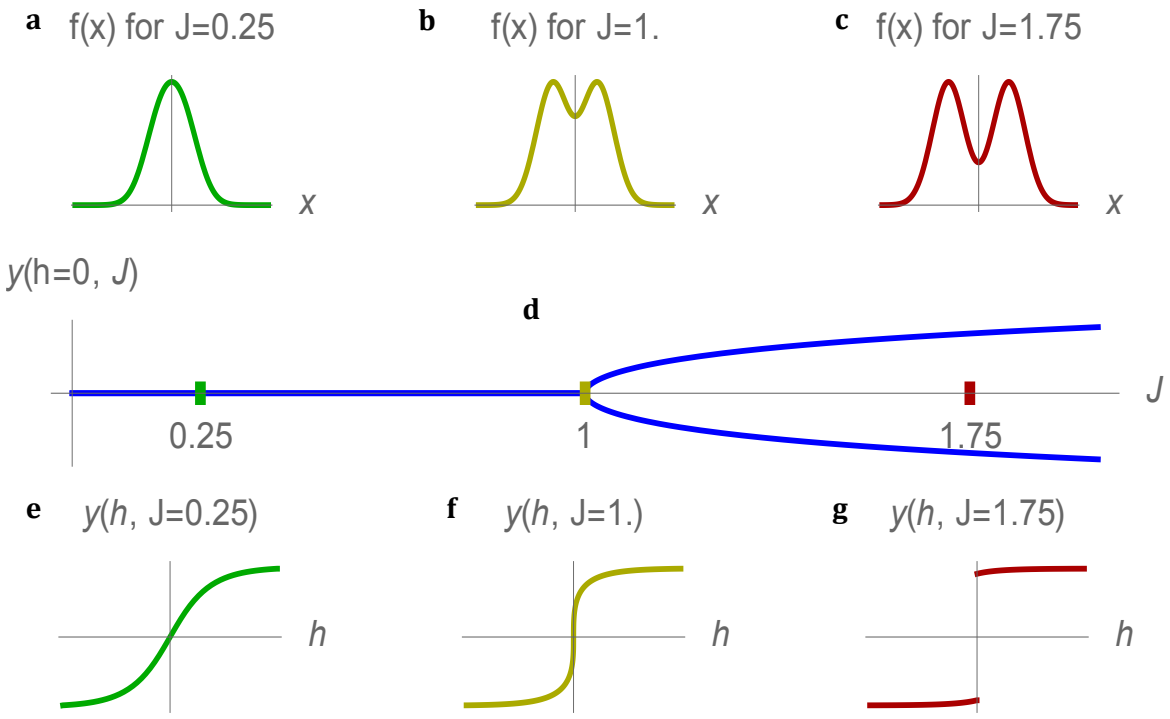


Figure 6-3: The stability of elections depends on the degree of electorate polarization. **a-c**, With increasing polarization (J) of the electorate opinion distribution, **d**, the electoral system undergoes a phase transition from possessing a single stable outcome to possessing two possible unstable outcomes. **e**, In the stable regime, the outcome smoothly responds to changes in the relative sizes of the two subpopulations (changes in h). **f**, At the phase transition ($J = 1$), the outcome is continuous but not differentiable in the relative sizes of the two subpopulations. **g**, In the unstable regime, the outcome discontinuously jumps. These figures were created using eqs. (6.10) and (6.11) for $a = \sigma = 1$ (for which $J = \Delta^2/2$).

6.3.3 Geographic considerations

These emergent discontinuities can couple to geospatial pattern formation. Given the existence of local elections, we can consider a spatially varying election outcome $y(\vec{r}_i)$, which serves as an order parameter for the $2D$ geographic system ($\vec{r}_i$ represents geographic location). Given local interactions, the $2D$ Ising model exhibits a universal behavior of spatial pattern formation [70]. Indeed, this is consistent with the observed geographic segregation associated with political polarization [265]. (As these dynamics are consistent with social influence [57] and homophily [243], we expect the order parameter $y(\vec{r}_i)$ to couple to these forces.) One such model that exhibits this universal behavior can be constructed by coarse-graining the local electoral outcomes $y(\vec{r}_i)$ into a statistical field [65, 66] $\tilde{y}(\vec{r})$, drawn from a statistical distribution with an effective local Hamiltonian (normalized by temperature) that allows for geographic heterogeneity:

$$\int d^2\vec{r}\big[t(\vec{r})\tilde{y}(\vec{r})^2 + u(\vec{r})\tilde{y}(\vec{r})^4 + K(\vec{r})(\nabla\tilde{y}(\vec{r}))^2 - h(\vec{r})\tilde{y}(\vec{r})\big] \tag{6.12}$$

where $u, K : \mathbb{R}^2 \to (0, \infty)$ and $t, h : \mathbb{R}^2 \to \mathbb{R}$. (Note that this Hamiltonian is intended to describe only the probability of various system configurations—not any of the dynamics.)

6.3.4 Political parties and primaries

We can also consider a model in which the political parties are explicit. Generically, the memberships of the two parties can be expected to differ in opinion due to the forces described in the previous paragraph. A model that captures this breaking of symmetry between the two parties can be described by the following temperature-normalized Hamiltonian for $p : \mathbb{R} \to [0, 1]$:

$$-\max_{y_1,y_2\in\mathbb{R}} \int_{-\infty}^{\infty} \beta f(x)[p(x)u_x(y_1) + (1 - p(x))u_x(y_2)]dx \tag{6.13}$$

where $p(x)$ denotes the probability that a citizen with opinion x belongs to or leans towards one of the two parties (with $1 - p(x)$ being the probability that they align with the other party), where β is a positive constant, and where the rest of the notation follows that of

eq. (6.7). As long as at least one of these parties chooses its nominee so as to maximize the probability of winning the general election (e.g. if primary voters vote based on this consideration), then no changes need be made to the analysis in eqs. (6.7) to (6.11). But, if *both* parties choose their nominees primarily based upon the opinions of their members rather than upon electability, then differences in the parties' members will generally cause the election to be unstable, regardless of general election voting behavior. For example, if each of the subpopulations in eq. (6.10) approximately corresponds to a political party, then nominees based on the median or mean opinions of the party members will be located at $\pm\Delta$, thus rendering the election unstable for all $\Delta > 0$ if the two parties have approximately equal support. Negative representation must then also be present; one such example would be if some opinions of the left-leaning party were to shift further to the left, resulting in a more left-leaning nominee that is less appealing to general election voters, increasing the likelihood that the right-leaning party's nominee wins.

More generally, this instability arises from primary electorates greatly differing from the general electorate. Given this inherently polarizing effect of the two-party system, electoral reforms such as instant-runoff/ranked-choice voting or approval voting that allow for third-party candidates to run without playing spoiler may reduce this instability. A full description of the stability of these voting systems, as well as parliamentary forms of government, is beyond the scope of this analysis and forms an important direction for future research.

6.4 U.S. presidential elections

Empirically, the opinions of the United States population, and in particular the opinions of those most likely to vote, have been polarizing over the past few decades [264], while voter turnout has remained approximately constant [240]. Thus, we might expect that over time, election outcomes have undergone a phase transition from a stable to unstable regime, and indeed this appears to be the case (fig. 6-4). (While elections cannot be expected to follow the precise assumptions behind eq. (6.11), they nonetheless may fall into the same *universality class* (section 6.6.8), including in the case where the space of opinions is multidimensional

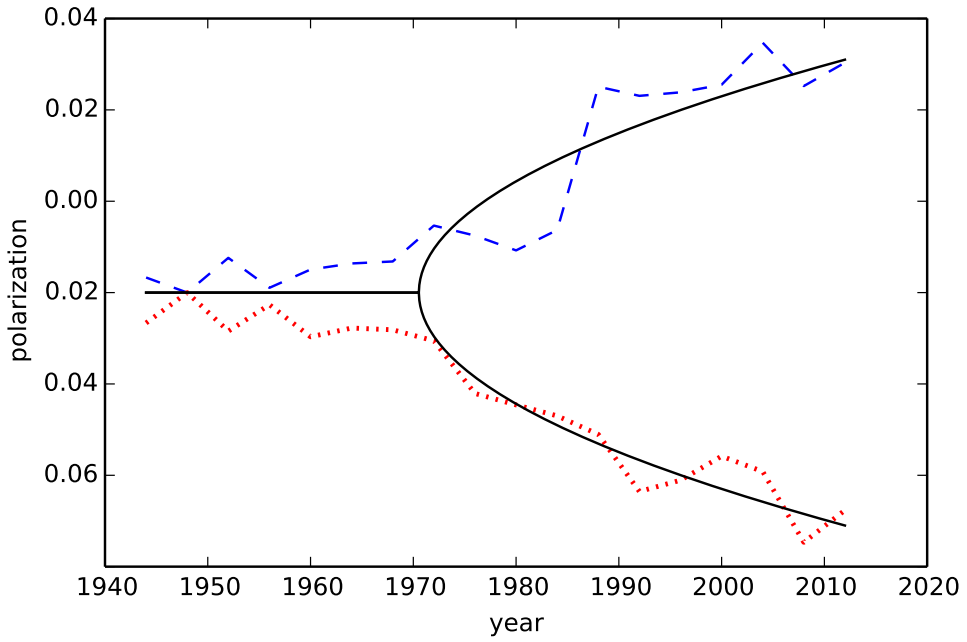


Figure 6-4: The polarization of the Democratic (dashed) and Republican (dotted) parties between 1944 and 2012, as measured by the fraction of polarizing words in the party platforms relative to a baseline. Party platforms are released once every four years. During the 1970s, there appears to be a divergence, which may correspond to a phase transition of the electoral dynamics into instability. The bifurcation associated with the universality class to which the mean-field Ising model belongs (solid, cf. fig. 6-3d) is superimposed; R^2 values are 0.86 for the Democratic party and 0.89 for the Republican party. See section 6.6.8 for further details.

(section 6.7.1).) Changes to the electoral process, such as reforms in the early 1970s as to how presidential party nominees are chosen, together with increasing polarization, may have driven this phase transition into instability. That the policies of legislators in politically homogeneous districts are more strongly correlated with the preferences of the district's median voter than the policies of legislators in heterogeneous districts [146] lends further empirical support.

6.5 Conclusion

In addition to whatever other problems may arise from instability, unstable elections also necessarily contain negatively represented opinions, a result that, due to the generality of the assumptions used to prove it, is expected to hold for any real-world election, regardless of the structure of the electoral process, the number of candidates, or other details. Therefore, the impact of electoral reforms on this instability should be considered. For instance, when voter turnout is low, political polarization can fundamentally shift electoral dynamics, causing large swings from election to election and leaving many negatively represented, regardless of the outcome. These results suggest that polices that increase voter turnout will not only result in more voices being heard but will also stabilize elections and reduce negative representation.

6.6 Methods

6.6.1 Translational invariance

In the main text, we make the assumption of translational invariance (eq. (6.1)). Technically, translational invariance is defined only in relation to a particular metric. Thus, the assumption of translational invariance can be relaxed without invalidating our results. For the proof that negative representation implies instability, all that is required is that the election be continuous (rather than invariant) under translations or formally, that for $f_c(x) = f(x + c)$,

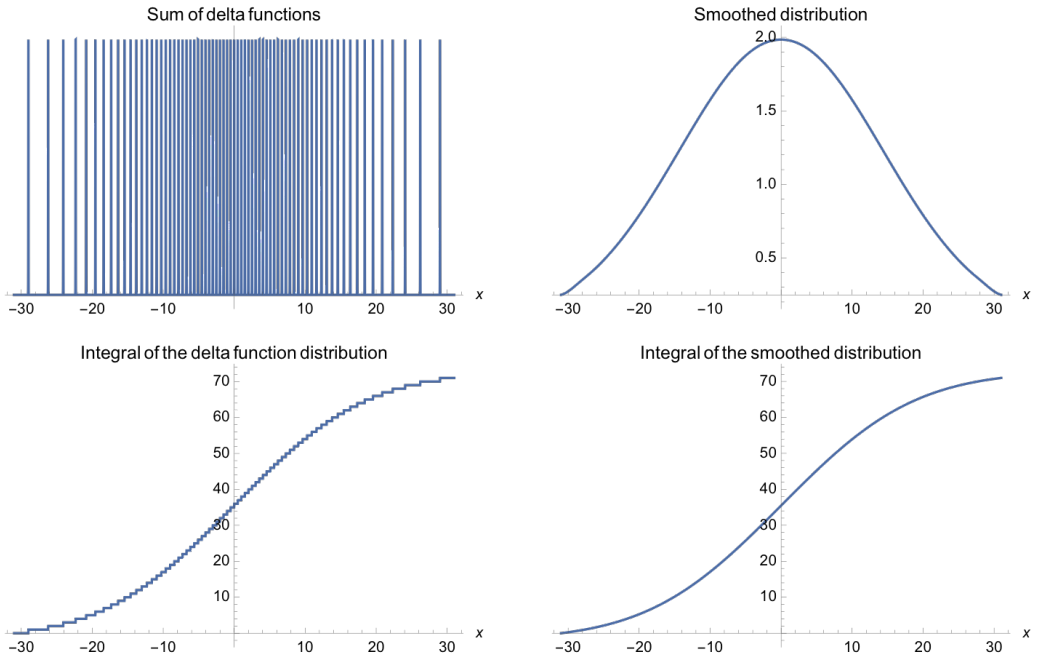


Figure 6-5: By replacing the Dirac delta functions with the approximation $\delta(x) \sim \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, a smooth distribution is obtained from the sum of delta functions. Note that the integrals of these distributions (taken from the lower bound of the domain of the graphs to x) are very similar, despite the striking differences between the distributions themselves.

$y[f_c]$ be continuous in c (note that this property is independent of the metric). The proof given in the main text then follows in the limit $\epsilon \to 0$. For the proof that the total representation sums to 1 (eq. (6.5)), there need only exist some metric on the opinion space under which the election is translationally invariant, so long as the representation is defined under such a metric.

6.6.2 Representation in the large-population limit

In this section, we derive properties of our representation measure when the number of citizens N is large. For a set of N citizens with opinions $\{x_1, x_2, ..., x_N\}$, the distribution of electorate opinions is $f(x) = \sum_{i=1}^{N} \delta(x - x_i)$, which is the only distribution satisfying the property that $\int_a^b f(x)dx$ is the number of citizens with opinions in the interval $[a, b]$. However,

it is often useful to choose $f(x)$ to be a smooth function that approximates $\sum_{i=1}^{N} \delta(x - x_i)$, in the sense that the difference between $\int_a^b f(x)dx$ and the number of citizens with opinions in the interval $[a, b]$ is no greater than 1 for all a, b (see fig. 6-5 for an example). For large enough N, the error of up to 1 opinion will generally not be significant. Alternatively, a natural interpretation of a smooth $f(x)$ is that the opinions are themselves probabilistic (for an example of explicitly probabilistic opinions, see section 6.7.2). Whether or not $f(x)$ is chosen to be smooth does not matter for the results of the text, although for the results that rely on the assumption that the number of citizens is large, the mathematics are simpler if $f(x)$ is assumed to be a function rather than a distribution. For instance, the expression for representation in the case of median voting involves evaluating f at its median, an operation which is not well-defined if f is a sum of Dirac delta functions.

In the limit of a large population ($N >> 1$), the change δf in the opinion distribution arising from an individual opinion will be small compared to the opinion distribution as a whole, and so we expand δy to first order in δf:

$$\delta y = \int_{-\infty}^{\infty} \delta f(z) \frac{\delta y}{\delta f(z)} dz \tag{6.14}$$

Note that eq. (6.14) does not apply to cases in which $y[f]$ is not differentiable, e.g. when the election is unstable and small changes in the opinion distribution can have an outsized impact; thus we do not use results derived from these equations when analyzing instability.

We now derive eq. (6.4). Note that when an individual opinion changes from x to $x' = x + c$, the opinion distribution changes by

$$\delta f(z) = \delta(z - x - c) - \delta(z - x) \tag{6.15}$$

where $\delta(z)$ is the Dirac delta function. Representation (eq. (6.2)) is then obtained in terms of functional derivatives of the election by substituting eq. (6.15) into eq. (6.14):

$$r_c(f, x) = \frac{1}{c}\delta y = \frac{1}{c}\left(\frac{\delta y}{\delta f(x + c)} - \frac{\delta y}{\delta f(x)}\right) \tag{6.16}$$

which, combined with eq. (6.3), yields eq. (6.4) (reproduced below).

$$r(f,x) = \frac{d}{dx}\frac{\delta y}{\delta f(x)} \tag{6.17}$$

By the fundamental theorem of calculus, we see from eq. (6.16) that $r_c(f,x)$ is the average of $r(f,x)$ over $[x, x+c]$.

We now prove eq. (6.5) ($\int_{-\infty}^{\infty} f(x)r(f,x)dx = 1$), which holds when the election ($y[f]$) is differentiable. For small ϵ, $f(z+\epsilon) = f(z) + \epsilon\frac{d}{dz}f(z) + O(\epsilon^2)$, and thus, using eq. (6.14),

$$y[f(z+\epsilon)] = y[f(z) + \epsilon\frac{d}{dz}f(z) + O(\epsilon^2)] = y[f(z)] + \epsilon\int_{-\infty}^{\infty}\frac{df(x)}{dx}\frac{\delta y}{\delta f(x)}dx + O(\epsilon^2) \tag{6.18}$$

Since f has compact support, which follows from f being an approximation of the opinions of a finite number of citizens, integrating by parts yields

$$y[f(z+\epsilon)] = y[f(z)] - \epsilon\int_{-\infty}^{\infty} f(x)r(f,x)dx + O(\epsilon^2) \tag{6.19}$$

which can be combined with eq. (6.1) in the limit $\epsilon \to 0$ to yield eq. (6.5). This proof assumes that f is differentiable for illustrative purposes, but the result will more generally hold if f and r are treated as distributions (generalized functions).

6.6.3 Nash equilibria of the electoral game

Consider a two-candidate election with endogenous candidacy: candidate positions (or, equivalently, candidates) are chosen in order to maximize the probability of victory. In this framework, the winner of the election will have adopted an unbeatable position y^*, provided such a position exists (i.e. a candidate with position y^* will have at least a 50% chance of winning against a candidate with any other position). Formally, (y^*, y^*) is a *Nash equilibrium*, since no candidate can improve her chances by changing her position. Because the voting game is symmetric, if (y_1, y_2) is a Nash equilibrium, then so are (y_1, y_1) and (y_2, y_2);

see, for instance, section 2.3 of [99]. Thus, if there is a unique Nash equilibrium, it must be of the form (y^*, y^*).

6.6.4 The utility difference model

Building upon the two-candidate election framework above (section 6.6.3), we describe the assumptions behind the utility difference model given by eq. (6.7) and give examples of its applications to median voting, mean voting, and an election between median and mean voting.

Denoting the probability that a vote from someone with opinion x will go to the first candidate minus the probability that it will go to the second by $p_x(y_1, y_2)$ (where y_1 and y_2 are the opinions of the first and second candidates, respectively), we assume there exists some function u_x such that

$$p_x(y_1, y_2) = u_x(y_1) - u_x(y_2) \tag{6.20}$$

This assumption yields an affine linear *utility difference model* [39], the potential election outcomes for which are given by eq. (6.7). Essentially, this model assumes that the position that maximizes a candidate's margin of victory does not depend on the position of the other candidate. We note that in order for eq. (6.20) to be able to be interpreted as a difference in probabilities, the functions $u_x(y)$ must satisfy

$$\max_{y_1, y_2} |u_x(y_1) - u_x(y_2)| \leq 1 \tag{6.21}$$

so that $|p_x(y_1, y_2)| \leq 1$ always holds.

For $u_x(y) = -a^2(y - x)^2$ where a is a positive constant, the mean opinion is selected, since for a random variable X, $\mathbb{E}[(X - \mu)^2]$ is minimized for $\mu = \mathbb{E}[X]$. (Here, y and the support of f must be confined to an interval of length at most a^{-1} in order to satisfy eq. (6.21).) For $u_x(y) = -a|y - x|$ (where again, a is a positive constant, and y and the support of f must be confined to an interval of length at most a^{-1}), the median opinion is

selected (since the median minimizes $\mathbb{E}[|X - m|]$), although by a different mechanism than the deterministic voting assumptions of the Median Voter Theorem [63]. (The Median Voter Theorem states that when political opinions lie in one dimension and everyone votes for the candidate whose opinion is closest to his, the optimal position for a candidate to take is that of the median voter.) Both of these functions can be viewed as limiting cases of the hyperbolic $u_x(y) = -\sqrt{a^2(y-x)^2 + b^2}$, with $b << a$ approximating median voting and $b >> a$ approximating mean voting. Under mean voting, for a citizen opinion either to the right of both candidates or to the left of both candidates, the farther away this opinion is, the stronger the citizen's preference between the two candidates. Under median voting, the strength of this citizen's preference for one candidate over the other is independent of how far the citizen's opinion is from both candidates. For the intermediate case, the strength of this citizen's preference gets stronger up to a point and then levels off as the citizen's opinion moves farther away from both candidates. However, in actual elections, citizens with opinions that are far from both candidates may be more likely to abstain from voting (or vote for a third-party candidate), which is why eq. (6.8) may be more realistic.

6.6.5 Representation in the utility difference model

In this section, we calculate representation for the utility difference model given by eq. (6.7). We derive the first part of eq. (6.9), and we then calculate representation for the examples given in section 6.6.4. To do so, we must assume there is a single possible election outcome y^* (see section 6.6.3). Then, from eq. (6.7),

$$y^* = \underset{\tilde{y}}{\operatorname{argmax}} \int_{-\infty}^{\infty} u_x(\tilde{y}) f(x) dx \tag{6.22}$$

which implies

$$0 = \int_{-\infty}^{\infty} u'_x(y^*) f(x) dx \tag{6.23}$$

Note that eq. (6.22) satisfies $y[f(x)] = y[\lambda f(x)]$ for any positive constant λ (scale invariance), and let $\tilde{f}(x) = f(x)/N$ where N is the size of the electorate, so that $\int_{-\infty}^{\infty} \tilde{f}(x) dx = 1$.

Considering the change that arises from the addition of a single individual with opinion x_0 to the population, we define δy^* by

$$y^* + \delta y^* = y[f(x) + \delta(x - x_0)] = y[\tilde{f}(x) + \epsilon\delta(x - x_0)] \tag{6.24}$$

where $\epsilon = 1/N$. Thus, substituting $y^* + \delta y^*$ for y^* and $\tilde{f}(x) + \epsilon\delta(x - x_0)$ for $f(x)$ into eq. (6.23),

$$0 = \int_{-\infty}^{\infty} u'_x(y^* + \delta y^*)(\tilde{f}(x) + \epsilon\delta(x - x_0))dx \tag{6.25}$$

Because $\int_{-\infty}^{\infty} u_x(x)f(x)dx$ is differentiable in f and has a single maximum in y, expanding eq. (6.25) to lowest order in ϵ yields

$$\delta y^* = \epsilon \frac{-u'_{x_0}(y^*)}{\int_{-\infty}^{\infty} u''_x(y^*)\tilde{f}(x)dx} + O(\epsilon^2) \tag{6.26}$$

Noting that the denominator is independent of x_0, of order 1 (i.e. independent of N), and negative (otherwise, y^* would be a minimum rather than a maximum),

$$\frac{\delta y}{\delta \tilde{f}(x_0)} = \lim_{\epsilon \to 0} \frac{\delta y^*}{\epsilon} \propto u'_x(y^*) \tag{6.27}$$

So, using eq. (6.4),

$$r(f, x) = \frac{d}{dx}\frac{\delta y}{\delta f(x)} = \frac{1}{N}\frac{d}{dx}\frac{\delta y}{\delta \tilde{f}(x)} \propto \frac{1}{N}\frac{d}{dx}u'_x(y^*) \tag{6.28}$$

If $u_x(y) = u(y - x)$, as it must be for some function u if the election is translationally invariant as in eq. (6.1), then

$$r(f, x) \propto -\frac{1}{N}u''(y^* - x) \tag{6.29}$$

Eq. 6.29 provides a direct link between citizen preferences and the representation of opinions. (If needed, the constant of proportionality can be determined through eq. (6.5).) Consider the examples for u given in section 6.6.4. For $u(y - x) = -a^2(y - x)^2$ and $u(y -$

$x) = -a|y - x|$, we can quickly derive the representation of opinions under mean voting ($r(f, x) = \frac{1}{N}$) and median voting ($r(f, x) = \frac{\delta(x-m)}{f(m)}$ where m is the median of f), respectively. For $u(y - x) = -\sqrt{a^2(y - x)^2 + b^2}$, which yields an outcome between that of median and mean voting,

$$r(f, x) \propto (1 + \frac{a^2}{b^2}(x - y[f])^2)^{-3/2} \tag{6.30}$$

resulting in the representation of opinions being concentrated around the election outcome, but not infinitely concentrated as it is for median voting. For u that are not concave, there will exist some x such that $u''(x - y) > 0$, and representation will be negative for those opinions (see eq. (6.29)).

6.6.6 Instability in the utility difference model

Here, we explore the conditions under which instability can arise in the model described by eq. (6.7), and we elaborate on the concrete model of instability with outcomes given by eq. (6.11). For the model given by eq. (6.7), $r(f, x) = \frac{d}{dx}\frac{\delta y}{\delta f(x)}$ is shown to be well-defined as long as $y[f]$ is single-valued, i.e. eq. (6.7) has a single maximum (section 6.6.5). Thus, instability can occur only when there are multiple maxima. (For a single maximum y^* with $\int_{-\infty}^{\infty} u''(x - y^*)f(x)dx = 0$, the functional derivative of y is not defined, but, as can be shown in a higher order analysis, there is no instability. In particular, for δy^* defined by eq. (6.24), we can derive

$$\frac{1}{6}(\delta y^*)^3 = \frac{-\epsilon u'_{x_0}(y^*)}{\int_{-\infty}^{\infty} \tilde{f}(x)u'''_x(y^*)dx} + O(\epsilon^{4/3})$$

in place of eq. (6.26), which yields $\delta y^* \propto \epsilon^{1/3}u'_{x_0}(y^*)^{1/3} + O(\epsilon^{2/3})$. Note that

$$r(f, x_0) = \frac{d}{dx}\Big|_{x=x_0} \delta y^*$$

is well-defined for any given $\epsilon = 1/N$—although it is not given by eq. (6.4) since $\frac{\delta y}{\delta f(x)}$ does not exist—thus, there is no instability.)

For an opinion distribution in which eq. (6.7) has two maxima y_1^* and y_2^*, if each position is taken by one candidate, then each candidate has a 50% chance of winning the election. However, an arbitrarily small change in the opinion distribution can favor one outcome over the other, giving either y_1^* or y_2^* a chance of winning that is arbitrarily close to 100% in the large-population limit. (For a finite population, the outcome of the election in this model is not deterministic, and the probability of a given candidate winning is continuous with respect to the opinion distribution. This discontinuity in $y[f]$ arising from multiple Nash equilibria in the large-population limit is analogous to a first-order phase transition.)

The existence of multiple maxima in y implies that $\int_{-\infty}^{\infty} u(x-y)f(x)dx$ is not concave (ignoring the degenerate case in which $\int_{-\infty}^{\infty} u(x-y)f(x)dx$ is constant over some interval). Thus, instability can arise only in the case of non-concave u, which is precisely the same condition under which negative representation occurs. That instability can arise only in the presence of negative representation should not surprise us, since it was proven under more general conditions in the main text. For this class of models, we also find that negative representation implies that there exist distributions of opinions for which instability arises, i.e. if u is not concave, then there exists an $f(x)$ such that eq. (6.7) has multiple maxima. To see why this is true, consider an opinion distribution f such that $f(x) = f(-x)$. Then, if there is a single maximum of eq. (6.7), it must lie at $y^* = 0$. In order for $y^* = 0$ to be a maximum, we must have $\int_{-\infty}^{\infty} f(x)u''(x)dx = 2\int_0^{\infty} f(x)u''(x)dx \leq 0$ (where the equality follows from the symmetry of u and f). But, assuming u is twice continuously differentiable and not concave, there exists an f such that $f(x) = f(-x)$ and $\int_0^{\infty} f(x)u''(x)dx > 0$. For such an f, $y^* = 0$ is not a maximum, thus contradicting our assumption that there was a unique maximum.

To provide an example of how, for non-concave u, the election is unstable for certain opinion distributions, we consider the u used for the example of negative representation: $u(y-x) = \exp\left[-\frac{(y-x)^2}{2a^2}\right]$ for some positive constant a (eq. (6.8)). We take the distribution of electorate opinions to be a sum of two (potentially unequally weighted) normal distributions

of equal variance:

$$f(x) = \sum_{\alpha=1,2} w_\alpha e^{-\frac{(x-\mu_\alpha)^2}{2\sigma^2}} \tag{6.31}$$

Then, from eq. (6.7), the outcome of the election is then given by

$$\underset{y}{\text{argmax}} \int_{-\infty}^{\infty} f(x) e^{-\frac{(y-x)^2}{2a^2}} dx = \underset{y}{\text{argmax}} \sum_{\alpha=1,2} w_\alpha e^{-\frac{(y-\mu_\alpha)^2}{2(a^2+\sigma^2)}} \tag{6.32}$$

Without loss of generality, we can assume that $\mu_2 = -\mu_1 \equiv \Delta \geq 0$. Defining the normalized election outcome $\hat{y} \equiv y/\Delta$, we solve eq. (6.32) to get the following condition:

$$\hat{y} = \tanh(J\hat{y} + h) \tag{6.33}$$

where $J = \Delta^2/(a^2+\sigma^2)$ and $h = \frac{1}{2}\ln\frac{w_2}{w_1}$. For $w_1 = w_2$ (i.e. for equally sized subpopulations), $h = 0$, and the election is stable with $\hat{y} = 0$ for $J \leq 1$ and unstable for $J > 1$. In the unstable regime, there are two possible outcomes described by $\pm|\hat{y}^*|$, and an arbitrarily small change in h can cause $\hat{y}$ to swing between its positive and negative value. In this regime, the majority of one of the subpopulations will be negatively represented: for $y^* < 0$, over half of the subpopulation centered at Δ will have opinions x with $x - y^* > \Delta - y^* > a$ ($\Delta > a$ in the unstable regime), and, from eq. (6.9), representation is negative for these opinions. Likewise, for $y^* > 0$ in the unstable regime, over half of the subpopulation centered at $-\Delta$ will be negatively represented.

6.6.7 Connection with the mean-field Ising model

We map the voting model that gives rise to eq. (6.33) (eq. (6.11) in the main text) onto a mean-field Ising model [192]. To begin constructing this map, note that the left-hand side of eq. (6.32) gives the limiting value of y as $N \to \infty$ when y is drawn from a probability distribution corresponding to the partition function

$$Z = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{-\frac{(y-x)^2}{2a^2}} f(x) dx \right]^N dy \tag{6.34}$$

Because y is a gaussian random variable, we can explicitly integrate over y, which yields, up to a multiplicative constant,

$$Z = \int_{-\infty}^{\infty} e^{-\sum_i \frac{(x_i-\bar{x})^2}{2a^2}} \Pi_{i=1}^N (f(x_i)dx_i) = \int_{-\infty}^{\infty} e^{-\frac{1}{2N}\sum_{i,j} \frac{(x_i-x_j)^2}{2a^2}} \Pi_{i=1}^N (f(x_i)dx_i) \tag{6.35}$$

This equation describes N interacting probabilistic "spins," with each spin weighted by the opinion distribution $f(x)$, with an energy penalty proportional to its mean-square distance from all of the other spins. For $f(x)$ given by eq. (6.31) (eq. (6.10) in the main text), the behavior is exactly that of a mean-field Ising model (with an external magnetic field for $w_1 \neq w_2$); in general, for bimodal symmetric $f(x)$ we expect a phase transition in which the system will spontaneously break the symmetry between the peaks as the peaks move farther apart. In the stable/disordered phase, both of the peaks of $f(x)$ are sampled by the "spins;" in the unstable/ordered phase, however, only one of the two peaks is sampled, and therefore the other peak is not represented. Despite the fact that each individual votes independently from everyone else, citizens are coupled through their collectively choosing a candidate, which is reflected by the effective interactions between the "spins." In the limit of weak interactions, i.e. $a \to \infty$, we recover mean voting, since in this limit, $u(y-x) = \exp\left[-\frac{(y-x)^2}{2a^2}\right] \approx 1 - \frac{1}{2a^2}(y-x)^2$. (Quadratic utility functions yield mean voting—see section 6.6.4.) Thus, mean voting is a way of "independently" aggregating opinions.

For a general $u(y-x_i) = \exp[-V(y-x_i)]$ (we can always write u in this form with $V(x) \geq c$ for some $c \in \mathbb{R}$), we note that eq. (6.7) yields an election outcome equivalent to the limit of y as $N \to \infty$ with y drawn from

$$Z = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{-V(y-x)} f(x) dx \right]^N dy = \int_{-\infty}^{\infty} e^{-\sum_i V(y-x_i)} \Pi_{i=1}^N (f(x_i)dx_i) \tag{6.36}$$

For quadratic V, we saw above that we could exactly integrate over the election outcome y to yield pairwise quadratic interactions between the x_i variables, but for general V, such an integration will yield many interaction terms of higher than quadratic order between these x_i. Although such integration cannot be carried out precisely, we expect this interacting system

to undergo a phase transition for bimodal $f(x)$ if the expansion of V produces sufficiently strong interactions. Thus, the system's behavior should be similar to the exactly solvable case in which V is quadratic.

6.6.8 Empirical data

To determine if the stability of U.S. presidential elections has changed over time, we used data from Jordan *et al.* [191] on the polarization in the party platforms. Jordan *et al.* used a combination of machine learning and human judgment to determine which of the frequently used words in the party platforms were polarizing and then determined the number of polarizing words (classified by political issue dimensions such as economic, foreign, etc.) in the Republican and Democratic platforms from 1944 to 2012. From this data, we calculated the total number of polarizing words as a percentage of all words in the platforms. We chose the percentage of polarizing words in the party platforms as a measure of political polarization over other measures of ideology—such as NOMINATE scores [271], which measure ideological purity based on agreement with other politicians—because we wanted an external, content-based measure of divergence in opinion rather than a measure of ideology that depends only on the positions that politicians take relative to one another. To construct fig. 6-4, we plotted by year the fraction of polarizing words in the Democratic platforms and the negative of the fraction of polarizing words in the Republican platforms. To correct for any time-independent bias affecting the number of polarizing words in the party platforms—for instance, which words Jordan *et al.* designated as polarizing—we subtracted from the data for each party separately the fraction of polarizing words from the year of least polarization for that party (which was 1948 for both parties). Thus, the data shown are the changes in the fraction of polarizing words relative to their baseline value (0.0258 for Democratic platforms and 0.0693 for Republican platforms).

As was noted in the main text and explained in section 6.6.7, our voting model (fig. 6-3) is equivalent to a mean-field Ising model. Real-world elections are unlikely to follow this model exactly, and even if they did, there is no reason to believe that there would be

a simple relationship between polarization (for which J is a dimensionless measure) and time. Nonetheless, if U.S. presidential elections underwent a phase transition in the same *universality class* as the mean-field Ising model, then in the vicinity of the phase transition, polarization would increase in proportion to the square root of the time from the transition, regardless of the precise relationship between time and polarization. In much the same way, magnetization increases near a ferromagnetic phase transition in proportion to $(T - T_c)^\beta$, where T is temperature, T_c is the temperature at which the phase transition occurs, and β is known as a *critical exponent*, which depends only on the universality class to which the phase transition belongs [193]. Inspired by this universality, we fit the polarization of both parties to the piecewise function

$$f(x) = \begin{cases} 0 & x \leq x_0 \\ A\sqrt{x - x_0} & x > x_0 \end{cases} \tag{6.37}$$

where x is the year and $f(x)$ is the fraction of polarizing words in that year's platform relative to the baseline value (see above), and A and x_0 are free parameters, corresponding to the amplitude of the polarization and the year that it begins, respectively. We found that $x_0 = 1970.54$ and $A = 0.0079196$ minimize the total sum of square errors, yielding R^2 values of 0.86 for the Democratic party and 0.89 for the Republican party. If the two parties are considered together, $R^2 = 0.87$.

6.7 Supplementary Text

6.7.1 Multidimensional opinion space

For the sake of simplicity, this chapter focuses on systems with a one-dimensional opinion space, but the concepts developed here can naturally be extended to a multidimensional opinion space, where the opinions of the electorate and candidates lie in $\mathbb{R}^n$, as in [99, 132,

39, 131]. This extension will be briefly outlined here. The definition of representation is generalized by replacing eq. (6.2) with

$$r_{\vec{c},\mu\nu}(f,x) = \frac{\delta y_\mu c_\nu}{|c|^2} \tag{6.38}$$

For a scalar measure, we use

$$\mathrm{tr}[r_{\vec{c}}] = \frac{\vec{c}\cdot\delta\vec{y}}{|c|^2} \tag{6.39}$$

When there exists an $r_{\mu\nu}(f,x)$ such that

$$\delta y_\mu = \sum_\nu r_{\mu\nu}(f,x)c_\nu + O(c^2) \tag{6.40}$$

for all $\vec{c}$, this $r_{\mu\nu}(f,x)$ can be used in place of eq. (6.3) as a representation independent of $\vec{c}$. In the large-population limit, eq. (6.16) is then replaced (using Einstein-summation notation) by the path-independent integral

$$\mathrm{tr}[r_{\vec{c}}(f,x)] = \frac{c_\mu}{|c|^2}\left(\frac{\delta y_\mu}{\delta f(\vec{x}+\vec{c})} - \frac{\delta y_\mu}{\delta f(\vec{x})}\right) = \frac{c_\mu}{|c|^2}\int_{\vec{x}}^{\vec{x}+\vec{c}} r_{\mu\nu}(f,x')dx'_\nu \tag{6.41}$$

where $r(f,x)$ (which satisfies eq. (6.40)) is a matrix defined by

$$r_{\mu\nu}(f,x) = \frac{\partial}{\partial x_\nu}\frac{\delta y_\mu}{\delta f(x)} \tag{6.42}$$

The differential representation in a direction given by the unit vector $\hat{v}$ is then given by $\hat{v}_\mu r_{\mu\nu}(f,x)\hat{v}_\nu$, which yields the same results as eq. 7 of [253] in the limit of a continuum of voters. The trace $\mathrm{tr}[r]$ gives a rotationally invariant scalar measure.

The representation normalization condition corresponding to eq. (6.5) is $\int f(x)r_{\mu\nu}(f,x)dx = \delta_{\mu\nu}$ where the integral is taken over $\mathbb{R}^n$.

In the multidimensional case, instability also implies a failure in representation. In a manner analogous to eq. (6.6), instability is characterized by

$$\lim_{\vec{c}\to 0} c_\nu r_{\vec{c},\mu\nu} \neq 0 \tag{6.43}$$

Generally, instability implies either that $\lim_{\vec{c}\to 0} |c| \operatorname{tr}[r_{\vec{c}}] \neq 0$, in which case negative representation (defined by $\operatorname{tr}[r_{\vec{c}}] < 0$) follows in the same way as the one-dimensional case, or that an infinitesimal change in opinion causes a finite orthogonal change in the outcome of the election. In this case, by considering further infinitesimal changes in opinion parallel to the first change in election outcome, and assuming that the magnitude of the change in the outcome of the election cannot grow without bound, one either gets negative representation directly ($\operatorname{tr}[r_{\vec{c}}] < 0$ for some $\vec{c}$)—or $\operatorname{tr}[r_{\vec{c}}] > 1$, from which negative representation follows as it does in the one-dimensional case.

Just as in the one-dimensional case, increasing polarization can drive the election through a phase transition from a stable to unstable regime, as eqs. (6.35) and (6.36) apply equally well if the opinions x_i are vector quantities. The nature of the unstable regime will depend on the symmetries present in the problem.[2] For example, consider a distribution of opinions given by

$$f(\vec{x}) = \sum_{\alpha=1}^{N_\alpha} e^{-\frac{(\vec{x}-\vec{\mu}_\alpha)^2}{2\sigma^2}} \tag{6.44}$$

where N_α is the number of subpopulations. For $N_\alpha = 2$, the instability that occurs for sufficiently large $|\vec{\mu}_1 - \vec{\mu}_2|$ will be very similar to what was described in the main text for the one-dimensional case. For $N_\alpha > 2$, if the set of points $\{\vec{\mu}_\alpha\}$ possess permutation symmetry, the system will resemble that of a mean-field Potts model (assuming the V in eq. (6.36) is isotropic). However, unlike the case for $N_\alpha = 2$—where regardless of the values of the $\vec{\mu}_\alpha$, the system possesses the appropriate symmetry to belong to the Ising universality class (due to the overall translational invariance of the problem and thus the ability to redefine the origin)—for $N_\alpha > 2$, we should not expect the system to generally possess the permutation

[2]However, as all of the possible phase transitions are described by a mean-field theory, they all belong to the same universality class.

symmetry necessary for Potts universality. Thus, we expect the real-world phase transitions that we observe to be of the $N_\alpha = 2$ type, i.e. of the same type of instability observed for a one-dimensional opinion space. For $N_\alpha > 2$, the situation may be able to be approximated by $N_\alpha = 2$, especially if there are dynamical social or political forces that pull the distribution of opinions towards a situation in which competition is roughly balanced between two opposing (potentially multi-party) coalitions at any given point in time.

6.7.2 The Owen-Shapley index as a special case

In this section we show that there exist functionals $y[f(x)]$ for which our representation measure (eq. (6.4)) reproduces the values of both the deterministic and probabilistic Owen-Shapley voting power indices. Thus, these voting power indices can be thought of as special cases of our measure. We give a brief background on the voting power literature and then consider the case of a one-dimensional opinion space, followed by a generalization to the case of a multidimensional opinion space for which the Owen-Shapley index was primarily designed.

When nothing is known about the preferences of voters, their political power has traditionally been measured by *a priori voting power* [136], which reflects the probability that a given individual or entity will cast the deciding vote and is usually measured by either the *Penrose index* [263] or the *Shapley-Shubik index* [295]. More precisely, to calculate a voter's *a priori* voting power, consider a random division of the rest of the voters into two camps. Then the probability that the excluded voter will get his way regardless of which camp he joins is his voting power; the Penrose index and the Shapley-Shubik index differ only in the way in which they randomly choose a division. While the Penrose index assumes that each voter randomly chooses one side or the other, the Shapley-Shubik index re-weights the probabilities so that each ordering of voters is equally likely.[3] These indices provide useful and counterintuitive results when the voters possess differing numbers of votes, as in

[3]There are some fundamental differences in the motivation behind the indices [136], but mathematically, they are rather similar, though neither is without drawbacks: while the assumptions behind the Penrose index are simpler, in general the sum of all voters' Penrose indices will not equal 1, while the sum of all voters' Shapley-Shubik indices will.

the European Union. For instance, under the 1958 voting rules for the European Economic Community, Luxembourg, despite having had one vote, had no voting power, since there were no possible divisions of the other five countries such that Luxembourg's vote would be decisive [144]. But in elections in which each voter has one vote, all measures of *a priori* voting power result in each voter having an equal amount of power. A measure of voting power that takes voter preferences into account is needed to determine how various opinions are differently represented. Many preference-based measures have been proposed—for instance, the spatial Shapley-Shubik index, also known as the Owen-Shapley index [258]—but, as we will see, these measures implicitly assume that people vote in a particular way.

In one dimension, the deterministic Owen-Shapley index [258] allows for only two possible orderings of the voters (left to right or right to left), for which the median voter is pivotal in both, thus yielding the same concentration of power that our representation measure (eq. (6.3)) yields in the case of median voting ($y[f] = \text{median}[f] \equiv m$ yields $r(f,x) = \frac{d}{dx}\frac{\delta y}{\delta f(x)} = \frac{\delta(x-m)}{f(m)}$). Benati and Marzetti [56] note that this extreme concentration of power is due to the deterministic nature of the Owen-Shapley model, which assigns zero probability to almost all orderings. They propose a generalized election model in which voters' opinions have both a deterministic and a random component. In the one-dimensional case, their treatment is equivalent to denoting the probabilistic opinion X_i of voter i by

$$X_i = x_i + \epsilon_i \tag{6.45}$$

where the x_i are deterministic and the ϵ_i are independent random variables with a continuous probability density function $f_\epsilon(\epsilon_i)$. Denoting the distribution of the x_i over the population by $f(x_i)$ (note that f will be a sum of delta functions for a finite population) and choosing an election in which people vote for the candidate closest to their probabilistic opinions X_i,[4] the Nash equilibrium for the two candidates' opinions is the median of the distribution

[4]This election takes the form of the *Random Utility Model* mentioned in section 2.2 of [39].

$f_X(X) \equiv (f * f_\epsilon)(x) = \int_{-\infty}^{\infty} f(x) f_\epsilon(X - x) dx$, i.e.

$$y[f] = \text{median}[f_X] = \text{median}[f * f_\epsilon] \equiv m \tag{6.46}$$

The continuity of f_X follows from that of f_ϵ, and we have made the additional assumption that $f_X(m) \neq 0$; otherwise, there is no unique Nash equilibrium.

We then calculate

$$\frac{\delta y[f]}{\delta f(x)} = \int_{-\infty}^{\infty} \frac{\delta \,\text{median}[f_X]}{\delta f_X(z)} \frac{\delta f_X(z)}{\delta f(x)} dz = \int_{-\infty}^{\infty} \frac{\text{sign}(z-m)}{2 f_X(m)} f_\epsilon(z-x) dz = \int_{-\infty}^{\infty} \frac{\text{sign}(z+x-m)}{2 f_X(m)} f_\epsilon(z) dz \tag{6.47}$$

which yields the representation measure (eq. (6.4)) for opinion i:

$$r(f, x_i) = \frac{d}{dx_i} \frac{\delta y[f]}{\delta f(x_i)} = \int_{-\infty}^{\infty} \frac{2\delta(z + x_i - m)}{2 f_X(m)} f_\epsilon(z) dz = \frac{1}{f_X(m)} f_\epsilon(m - x_i) \tag{6.48}$$

The rightmost side of eq. (6.48) is the probability that voter i is the median—i.e. pivotal—voter; thus, $r(f, x_i)$ is equal to the generalized Owen-Shapley index for voter i.

The Owen-Shapley index was developed primarily for multidimensional opinion spaces in $\mathbb{R}^n$. Owen and Shapley [258] consider a randomly drawn unit vector $\hat{v} \in \mathbb{R}^n$ and then order individuals by defining $i < j$ if $\vec{X}_i \cdot \hat{v} < \vec{X}_j \cdot \hat{v}$. (The $\vec{X}_i$ are deterministic but can easily be modified to be partially probabilistic as in [56].) The Owen-Shapley index of i is again the probability that i is the median of the resulting ordering. To see how this power index is a special case of our multidimensional representation measure (eq. (6.42)), consider the following method of choosing a candidate, given a set of voters with (potentially probabilistic) opinions $\vec{X}_i \in \mathbb{R}^n$:

1) Randomly choose an orthonormal basis $V = \{\hat{v}_1, ..., \hat{v}_n\}$ for $\mathbb{R}^n$.

2) Conditioning on the orthonormal basis V, let $m_\alpha(V)$ be the median of the probability distribution function for $\hat{v}_\alpha \cdot \vec{X}$, where $\vec{X}$ is randomly drawn from the voter opinions $\vec{X}_i$.

3) The election outcome is then given by $\vec{y}[f] = \sum_{\alpha=1}^{n} m_\alpha(V) \hat{v}_\alpha$.

Note that $\vec{y}$ is now a random variable (since it depends on the orthonormal basis V), and so the right-hand side of eq. (6.42) must be replaced by its expectation value, i.e. $r_{\mu\nu}(f,\vec{x}) = \mathbb{E}[\frac{\partial}{\partial x_\nu}\frac{\delta y_\mu}{\delta f(\vec{x})}]$. From eq. (6.48) (with $\hat{v}\cdot\vec{y}$, $\hat{v}\cdot\vec{x_i}$, and $\hat{v}\cdot\vec{\epsilon_i}$ substituting for y, x_i, and ϵ_i), $\hat{v}_\mu r_{\mu\nu}(f,\vec{x}_i)\hat{v}_\nu$ is equal to the probability that i will be the median voter along $\hat{v}$. Therefore, $\mathrm{tr}[r(f,\vec{x}_i)]$ is equal to the expected number of basis vectors along which i will be the median, and so $\mathrm{tr}[r(f,\vec{x}_i)]$ is equivalent to n times the Owen-Shapley index.

The agreement between $r(f,\vec{x})$ and the Owen-Shapley index follows from the fact that $\frac{d}{d(\hat{v}\cdot\vec{x}_i)}\frac{\delta\,\mathrm{median}[f_{\hat{v}\cdot\vec{X}}]}{\delta f(\vec{x}_i)}$ measures the probability that i is the median voter along $\hat{v}$ for this class of voting models. In this sense, the Owen-Shapley index of power implicitly assumes an election in which some sort of median is chosen. This model is appropriate when voters vote deterministically (although the options presented for them to vote on may be random). But such deterministic voting assumes that voters distinguish between very small differences in policy with 100% certainty, and it also assumes that there is no chance that a voter abstains. While these assumptions may hold for assemblies of elected officials (and in particular to the EU, where these measures are most commonly applied), they tend to fail for mass elections, in which a citizen may sometimes vote for the candidate farther from her opinion and sometimes may choose not to vote at all.

6.7.3 A brief review of social choice theory

Here, we briefly review the social choice literature on elections, focusing in particular on the aspects of spatial models of elections—i.e. models that denote the political preferences of a citizen by a point in some space that can be embedded in $\mathbb{R}^n$ for some n—that are most relevant to our work. (These points are often referred to as *ideal points*—rather than *opinions*, as in this chapter—in order to emphasize the simplification that takes place when opinions are placed in a low-dimensional space.) The social choice literature also considers more general questions concerning the aggregation of opinions, often with counterintuitive results [114, 20, 150, 284, 128];[5] see [223] for an overview. The simplest spatial models of elec-

[5]For example, Arrow's Impossibility Theorem [19] states, roughly speaking, that a dictatorship is the only system of aggregating preferences that 1) prefers A to B if the citizens unanimously prefer A to B,

tions take $n = 1$, i.e. they assume that political preferences can be determined by a citizen's position on a line (often interpreted as the left-right political spectrum), e.g. [177, 63, 123]. These models assume that there are two candidates competing for a majority of the votes and that all citizens vote and that they vote for whichever candidate is closest to them. Under these conditions, there is a pure Nash equilibrium (see Methods section 'Nash equilibria of the electoral game') in the candidate strategies, which is for both candidates to adopt the position of the median voter. Downs [123] recognized the limitations of these assumptions and intuited qualitatively that due to the nonvoting that can arise from alienation, "democracy does not lead to effective, stable government when the electorate is polarized."

Subsequent work has focused on the existence of Nash equilibria in candidate strategies under various assumptions concerning voter and candidate motivations. In a multidimensional opinion space ($n > 1$), the deterministic voting that gives rise to the Median Voter Theorem for $n = 1$ does not yield a Nash equilibrium without additional restrictions on the voters' preferences (see e.g. [269, 113, 112]). Nash equilibria can often be restored by considering spatial models in which citizens cast votes probabilistically, since probabilistic voting allows for smoother changes in voting behavior with respect to changes in the candidate positions (e.g. [182, 99, 132, 219, 100]; see [39] for a review). Other assumptions include, for instance, voter abstention [175, 174], policy-motivated candidates (e.g. [80, 220]) and game-theoretic considerations on the part of the voters (e.g. [211, 242]).

In our work, we are concerned with the the functional $y[f(x)]$, which maps the distribution of citizen opinions onto the electoral outcome, rather than with the mechanism that gives rise to this map. In the latter half of the main text, we build upon the probabilistic voting models developed in the literature, relaxing the assumption of concave voter preferences so as to provide a concrete demonstration of how negative representation and electoral instability can arise, the latter of which is reflected in the existence of *multiple* Nash equilibria.

and 2) does not allow its preference between two alternatives A and B to be affected by citizens preferences concerning a third alternative C.

Chapter 7

Geospatial Political Polarization

A key aspect of complex systems is their multi-scale structure (Chapter 1). In this chapter we consider the multi-scale structure of political opinion and some of its implications for multi-scale governance.

7.1 Introduction

In recent years, the American electorate has become increasingly polarized [266], prompting both concern and research into the potential mechanisms behind this change. Such concern is justified: the more polarized the electorate, the more difficult it is for politicians to represent that electorate in any meaningful way [123]. Furthermore, if political divisions overlap with divisions among social networks, as appears to be the case [179], individuals may feel that not only they, but also their entire social groups, are not represented, thereby deepening social and political divisions.

Researchers have found many mechanisms that contribute to polarization; these mechanisms may be classified into two categories: "individual" and "group." Individual polarization occurs independently of social contacts. For instance, many have argued that the proliferation of news sources has enabled individuals to consume news that reinforces their preconceived beliefs [185, 272]. Furthermore, individuals consuming even the same media

may interpret the information differently, in accordance with their preconceptions [362, 225]. These polarization mechanisms are individual in that they do not depend on whom a person knows.

Group polarization, by contrast, refers to the well-documented phenomenon in which the average opinion of a group of individuals with similar opinions grows more extreme following deliberation, in some cases becoming more extreme than any of the original individual opinions [183]. This tendency towards extremism is thought to be the product of information exchange and social pressure [315]. Information exchange refers to the effect of social contacts on the pool of information available to an individual. Social pressure arises as individuals gain confidence in their views as a consequence of their social contacts affirming their opinions [57, 204]. If a social group encompasses a healthy mix of opinions, social interactions serve to moderate the individual opinions of group members [124]. But, by facilitating the selection of like-minded political discussion partners, the Internet and increasing geographic mobility may result in more homogeneous groups whose views then become more extreme over time.

The above discussion reviews how individual and group behaviors give rise to polarization. However, national political opinion cannot be fully understood in terms of individual- and group-level behaviors alone because the global properties of complex systems—such as our society—are determined by the interactions *between* groups. Therefore, this chapter will attempt to address this issue from a different perspective: it will focus on how the *universal* properties of systems—the large scale patterns that emerge regardless of small-scale details—can be used to understand certain aspects of polarization and form an avenue for further research.

Here, we consider geographic patterns rather than on particular social mechanisms. Most (though not all) treatments of public opinion do not consider the effect of geographic location, or do so merely to note that geographic differences exist. These treatments view individuals as either independent—interacting with nobody—or subject to general social forces, inter-

acting with others regardless of their location. This is insufficient: due to spatial correlations in political opinion (section 7.3.1), the effects of location extend beyond simple differences.

7.2 Theory

7.2.1 A Gaussian field theory

Some of the assumptions behind the theory presented in this section (section 7.2.1) and associated sections in the appendix are later shown to be unlikely to hold (see section 7.3.2). (Note that the assumptions in question are not used in section 7.2.2; this disclaimer applies only to this section.) I nonetheless include what follows for three reasons. First, the analysis is instructive and, even if it does not apply to the situation at hand, could still be used in other contexts. Second, some insight about the system being modeled can nonetheless be drawn, especially if causal claims are replaced with descriptive ones. And third, I believe it to be illustrative to have a record of the theory as it was developed rather than a retrospectively modified version.

This model recognizes the futility of attempting to capture individual behavior and instead considers, at each geographic location, the aggregate political opinion of people within a certain radius of the location.[1] Although the behavior of individuals is impossible to predict, aspects of their collective behavior can be elucidated (see section 1.2). So while this model will inevitably fail to capture all of the local forces that give rise to polarization, it will reveal the universal properties of polarization—those that do not depend upon the details of the system—just as the existence of sound waves does not depend on material composition.

[1]Formally, people occupy a geographic position $x \in \mathbb{R}^2$. Their opinions lie in some multi-dimensional space $\mathbb{R}^n$ (as in the spatial model of elections [123, 99, 132]) with the average opinion of the people living in some radius r around location x denoted by $m(x)$. This radius r must be chosen to be large enough such that $m(x)$ is smoothly varying, and the parameters of the model must be implicit functions of r so that resulting large-scale behavior is independent of the precise choice of r. If one wishes to account for varying population densities, one can choose r to differ based on the local population density, such that $m(x)$ corresponds to the aggregated opinions of the same number of people for all x.

In this model, the aggregate opinion of a locale is influenced by intrinsic factors, which arise from within the locale, and by interactions with the aggregate opinions of other locales.[2] The effect of the intrinsic factors on a particular locale is modeled by an intrinsic parameter for that locale. Were there no interactions with other locales, the aggregate opinion of a locale would simply equal the locale's intrinsic parameter, plus some statistical noise. Because political opinion is influenced by social connections and because one's social connections are likely to live nearby, inter-locale interactions make it more likely that nearby locales have similar aggregate opinions. These inter-locale interactions are modeled by a single parameter that captures the average influence of nearby social connections on political opinion. Given this social influence parameter and the locales' intrinsic parameters as inputs, this model outputs a statistical distribution for the aggregate political opinion at all geographic locations.

As it turns out, this model is exactly solvable,[3] allowing a wide range of behaviors to be explored. For instance, some have claimed that America has been geographically self-sorting over the past few decades, due to the fact that Americans move far more often than they used to, coupled with the (often subconscious) tendency to move to areas inhabited by like-minded individuals [62, 316, 189]. Even in this simple model, such geographic sorting greatly exacerbates polarization (see figures in section 7.5.2). Before geographic sorting, the intrinsic forces that drive differences in political opinion are largely suppressed by the different intrinsic forces of nearby locales; but after sorting, such damping is diminished and intrinsic differences are unchecked.[4]

---

[2]Mathematically, the aggregated political opinion at each point in space, $m(x)$, is treated statistically, drawn from the probability distribution $\mathbb{P}[m(x)] = \frac{1}{Z} e^{-f[m(x)]}$ with $f[m(x)] = \int d^2x \left[\left\{ \frac{1}{2} m(x)^2 - h(x)m(x) \right\} + \frac{K}{2} (\nabla m(x))^2 \right]$ (where Z is a normalization constant). We can think of $f[m(x)]$ as proportional to the "energy" of the configuration $m(x)$, configurations of higher energy being exponentially less likely to occur. The first two terms in $f[m(x)]$ represent the intrinsic forces, while the third represents the interactions with nearby locales. See section 7.5.1 for more details.

[3]Performing the Fourier transforms $m(x) = \frac{1}{A} \sum_q \tilde{m}(q) e^{iqx}$ and $h(x) = \frac{1}{A} \sum_q \tilde{h}(q) e^{iqx}$ (with $A = \int d^2x$ being area of the system), we obtain the expected values of the opinion Fourier modes: $\langle \tilde{m}(q) \rangle = \frac{h(q)}{1+Kq^2}$

[4]More rigorously: when translating these intrinsic forces into political opinion, $\tilde{h}(q)$ is decreased by a factor of $1 + Kq^2$. Prior to self-sorting, we expect the intrinsic forces to be uncorrelated for nearby points in space, and so $\tilde{h}(q)$ will be dominated by rapidly varying (high-q) Fourier modes, but after self-sorting, $\tilde{h}(q)$ will be dominated by more slowly varying (lower-q) Fourier modes, which are suppressed by a smaller factor.

It should be noted that polarization arises in this model from the geographic clustering of similar forces, regardless of what the cause of this clustering is. Thus, this model explains how polarization is exacerbated not only by geographic sorting, but by *any* force that divides America across geographic lines. Such forces include the increasing alignment of political parties along urban-rural lines, as well as any economic, cultural, religious, or political shifts that increase (or make more salient) the differences between city and country living.

Another implication of this model is that a decrease in the strength of local social influences results in greater polarization.[5] Such a decrease can have a variety of causes, including a decrease in social capital [273] and a displacement of local social connections by long-distance ones.

7.2.2 Geographic scale

An important aspect of this model is the scale over which opinion is aggregated. The opinion at any given location may be averaged over those living within a few miles, or hundreds of miles, or any distance in between. Depending on the distance over which opinion is aggregated, political opinion can look quite different.[6] In particular, it is possible to look at political opinion with fine granularity and see a polarized distribution of opinions, but when one decreases the granularity by aggregating over larger and larger geographic regions, one is left with regions that are on average more moderate (section 7.5.6).

But why is scale important? If individuals diverge strongly in opinion, does it matter whether or not aggregating opinion over a larger scale makes the distribution more moderate? In short, yes. Divergence of opinion in and of itself does not make for an unhealthy political climate; it is only when this divergence is accompanied by segregation and tribalism—when people do not talk to those with opposing views, and when these homogeneous groups have a tribal us vs. them mentality towards those with whom they disagree. The larger the distance over which opinion must be aggregated until most regions are moderate, the more

[5]The parameter K represents the average strength of local social influences. A smaller K leads to larger variations in opinion (i.e. larger values of $\tilde{m}(q)$ for $q \neq 0$).

[6]Recall from the first footnote that the parameters of the model depend on the radius r over which opinions are averaged to give the opinion $m(x)$ as a continuous function of location x.

readily such divergences in opinion lead to a tribal and segregated political climate, due to the geographic nature of both social ties and multi-scale government (i.e. politicians representing geographically localized groups). If individuals differ in opinion but the average political opinion of towns and cities is moderate, local politics may be contentious, but we can expect a rather moderate state and national political atmosphere. If towns and cities have divergent opinions, but the aggregate opinions of states are moderate, we can perhaps expect contentious politics at the state-level, but national politics should remain moderate. If, when aggregated at the state level (or even larger scales), opinion is still polarized, then we should expect to see a tribal Congress and presidential elections that divide rather than unite.

This analysis highlights the importance of considering geographic variation in opinion, since the distribution of opinion depends on the scale at which it is observed. At the smallest scales, we should always expect to find some degree of polarization, for that is human nature; the relevant question is not whether society is polarized but rather to what scale does polarization persist.

In order to capture the effects that technologies such as the Internet have had on political opinion, long-distance interactions can be included.[7] It can be shown that these connections will, as compared to local connections, result in more ideological homogeneity within political tribes and consequently less common ground between them. Furthermore, if long-distance interactions are self-selected—such that individuals with similar opinions are more likely to interact—rather than random, then the populace can spontaneously fragment into two opposing tribes, even in the absence of intrinsic differences.[8] This spontaneous fragmentation cannot occur when interactions are predominantly local (even if they are self-selected), which suggests a potential key difference between the polarization of today and that of the past.

[7]This can be done by including the term $\int \frac{d^2x d^2y}{2A} J(m_x - m_y)^2$ in $f[m]$.

[8]Such self-selected long-ranged interactions can be captured by the term $\int \frac{d^2x d^2y}{2A} J(m_x - m_y)^2 m_x m_y$. For certain parameter regimes, the presence of this term makes the most likely configuration of opinions one in which opinions are clustered around opposite poles (see section 7.5.7 for details).

7.3 Precinct-level voting data from the 2016 U.S. presidential election

7.3.1 Spatial structure

Here we examine in a series of figures the geospatial distribution of political opinion. Our analysis in this section is acausal; we simply examine the geospatial correlations in political opinion. In addition, we consider the geospatial correlations in political opinions as predicted by demographic variables, as well as the geospatial correlations in the residuals from these predictions.

The fact that the residuals have geospatial correlations to the extent that they do is indicative of there being factors not explained by the demographic variables we considered (age, sex, race, income, and educational level) that are substantially geographically correlated and had a predictive effect on the 2016 outcome. Various regions at various scales are either more conservative or more liberal than would be predicted by demographics alone in a fractal-like way. In other words, even after accounting for demographics, there are geographically correlated factors that cause same regions (e.g. the San Francisco Bay Area, Vermont, etc.) to be substantially more liberal than expected, while other areas are substantially more conservative.

7.3.2 Neighbor effects

In this section, we analyze spatial voting and demographic patterns with the goal of determining whether or not the political preferences of one's geographic neighbors (specifically, those living in nearby precincts) have any *causal* effects on one's vote.

Set-up

We analyzed data from the 2016 U.S. presidential general election. We define the *Republican two-party vote-share* (henceforth known simply as the vote-share) of a precinct as the number

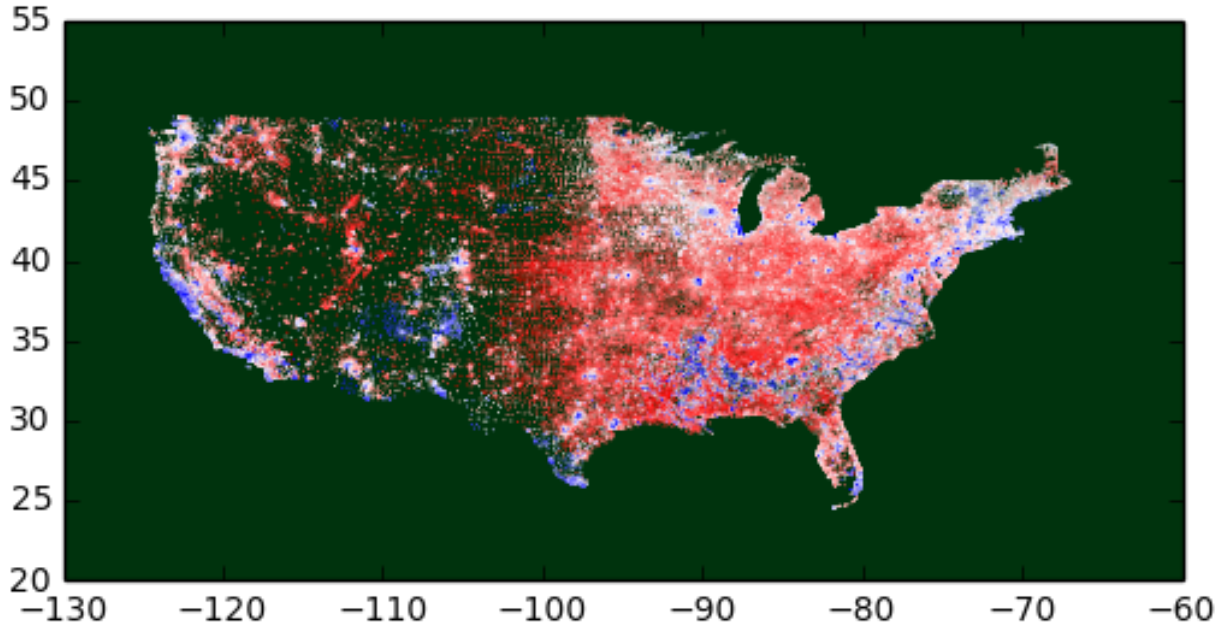


Figure 7-1: The share of the two-party vote that Trump won in 2016 for each precinct. Each dot is a precinct, with its color denoting Trump's vote share; the scale runs from blue (0%) to white (50%) to red (100%).

of votes in that precinct for the Republican nominee (Trump) divided by the the total number of votes for either major party (Republican or Democratic) nominee (Trump or Clinton). Note that the vote-share will be a rational number between 0 and 1, inclusive, and that it will not be defined for precincts in which neither Trump nor Clinton received a single vote. A precinct's *demographically predicted vote-share* is the best estimate of that precinct's vote share based on the distributions of age and sex, race, income level, and educational attainment within that precinct).

In order to analyze causal effects, we asked the following question: do the demographic variables from neighboring precincts explain some of the variance in a precinct's vote share that is not explained by the demographic variables from that precinct? In other words, can the demographically predicted vote-share of a precinct be substantially improved by the demographically predicted vote-shares of its neighboring precincts?

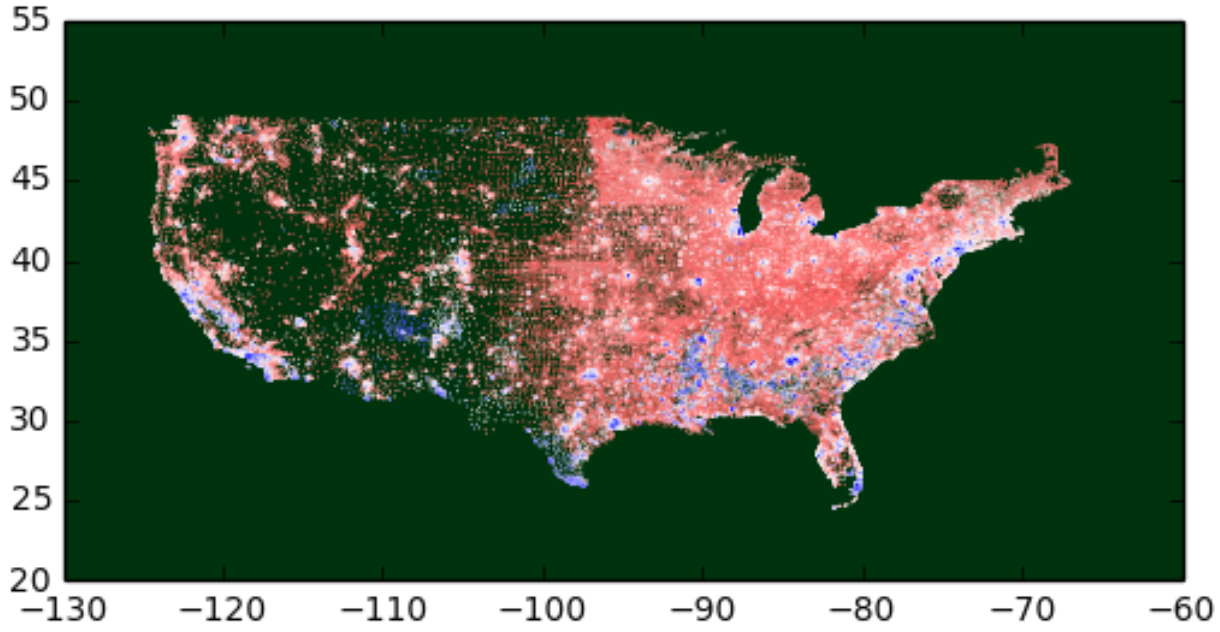


Figure 7-2: The predicted share of the two-party vote for Trump in each precinct, using a random forest model based on demographic information. The color scale is the same as in fig. 7-1.

We note that if the answer to this question were yes, it would suggest but not prove such causal influence. (Such correlations could be acausal if there were some factor independent of a precinct's own demographics but that was somehow correlated both with its vote-share and with its neighboring precinct's demographics.)

However, if the answer to this question is no, then any causal impact of a precinct's neighbors on its vote-share would have to be completely uncorrelated from the neighboring precincts' age, sex, race, income, and educational attainment distributions. In theory, one could imagine causal pathways in which only the aspects of political opinion that were uncorrelated with these demographics had any causal influence, or, equivalently, that the causal influence of the aspects of neighboring precincts' political opinion that was due to demographics was precisely canceled by other effects of their demographics or factors correlated with their demographics. In practice, such a precise cancellation seems implausible.

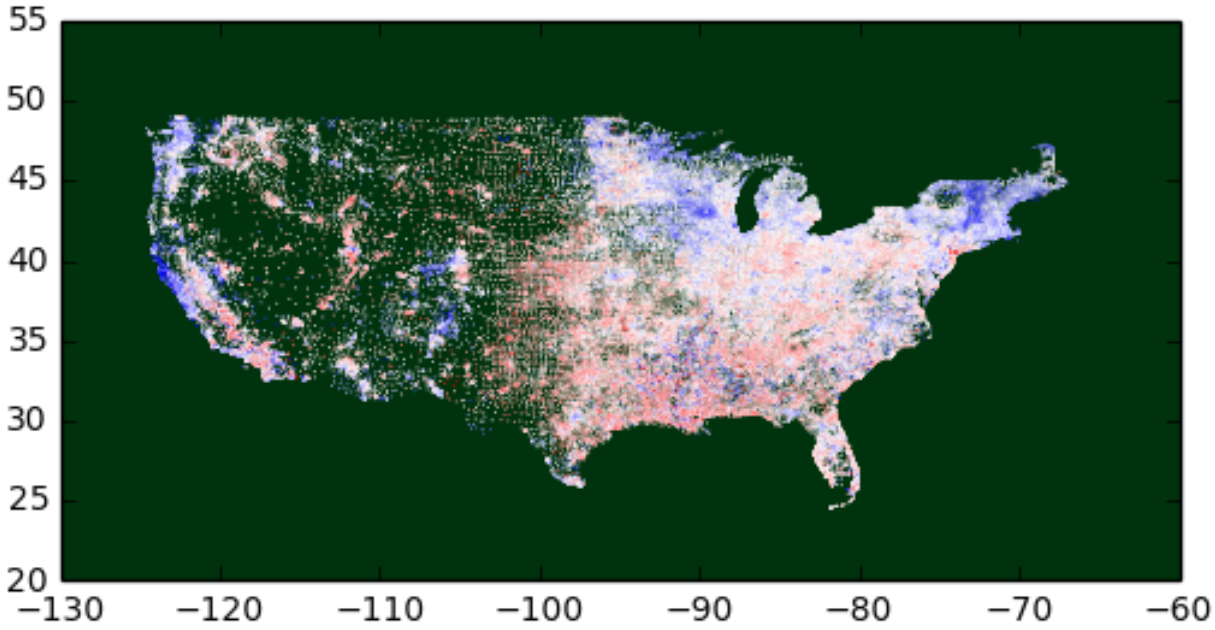


Figure 7-3: The difference between the actual outcome and the random forest prediction based on demographics (blue=outcome more Democratic than predicted; red=outcome more Republican than predicted). The color scale runs from blue (residual ≤ -0.5) to white (residual $= 0$) to red (residual ≥ 0.5).

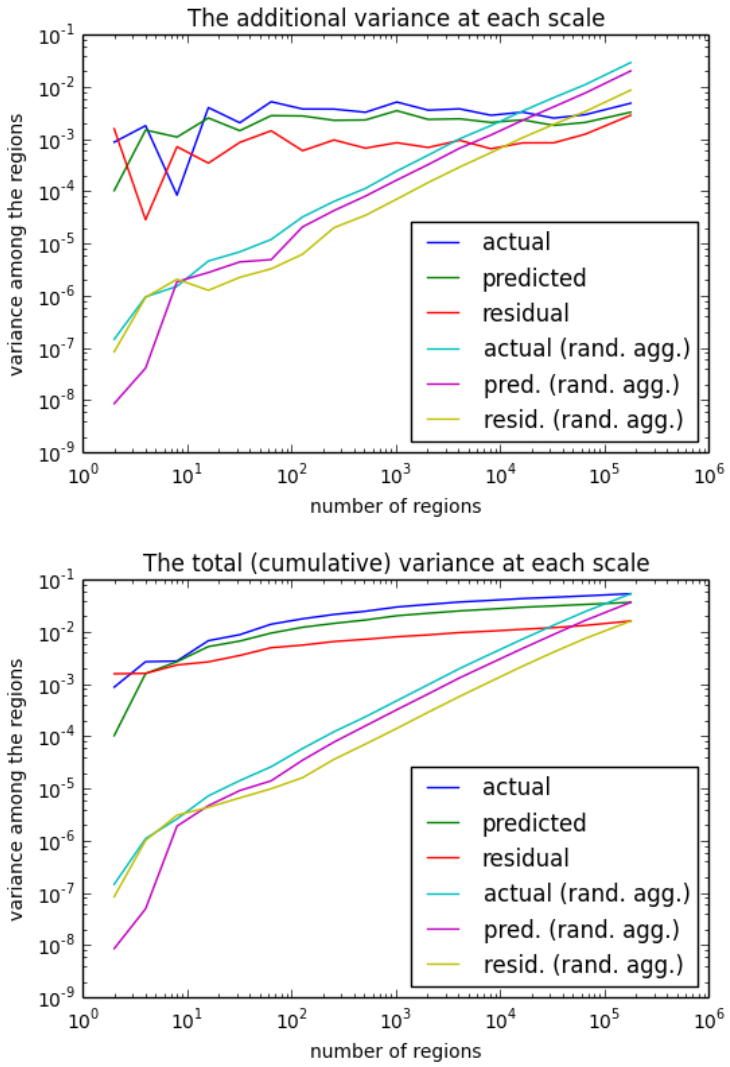


Figure 7-4: The variance in the {actual (blue) / predicted (green) / residual (red)} Republican share of the two-party vote at each level of geographic aggregation as a function of the number of regions that result from the aggregation. As aggregation occurs and smaller regions are aggregated into larger ones, the number of regions decreases. The cyan, yellow, and magenta curves indicate the same, except if precincts are aggregated randomly rather than with their geographic neighbors. The right-most data points indicate the variance of the values for precincts (of which there are nearly 200,000), while the left-most data points indicate the variance between the means of the two groups in to which all precincts have been aggregated (for the geographic aggregation, that was performed using a k-d tree, the two groups are obtained by dividing the regions according to whether they lie to the north or south of the median latitude of the precincts). Note that for random aggregation the total variance is proportional to the number of regions (deviations from a line with slope 1 are due to idiosyncrasies in the particular random aggregation shown—averaging over all random aggregations would lead to a perfectly straight line). The variances for geographic aggregation decrease far more slowly, indicating substantial geographic correlation. Note also that for the geographic aggregation the variance decreases in an approximately scale-free fashion (i.e. linearly on a log-log scale) or alternatively note that the additional variance that arises as one moves to finer and finer scales is roughly constant, indicating that there is substantial variance in opinion across all scales.

It should be noted that regardless of whether or not a precinct's neighbors have causal influence over its vote-share, we should expect the residuals of the predictions of vote-shares from demographics to be correlated among nearby precincts, because there are other factors besides the demographic variables considered that are spatially correlated and that influence vote-share. Indeed, such patterns were observed in section 7.3.1.

Results

We find that, conditioning on a precinct's demographic variables, including the vote-share prediction based on demographic variables for neighboring precincts does not significantly improve the predictive power of the model. As explained above, this finding suggests a lack of any significant direct causal influence. If they did, we would, contrary to our observation, expect precincts to be more liberal (conservative) than expected if their neighboring precincts had more liberal- (conservative-) leaning demographics (unless such causal forces were somehow orthogonal to correlations in each precinct's neighbors' age, sex, race, income, and educational attainment distributions).

These results imply that when considering the theory in this chapter, the parameter K that is responsible for spatial correlations in political opinion (see sections 7.5.1 and 7.5.6) cannot be straightforwardly interpreted as an effect of social influence. Clearly, however, such spatial correlations in political opinion exist (section 7.3.1), and so the inclusion of the parameter K is not inappropriate, but care must be taken when considering which real-world factors enter the model through K. For instance, the causal arguments presented in sections 7.2.1 and 7.5.2 that a spatial self-sorting of political opinion contributed to polarization due to individuals being less moderated by their geographic neighbors is unlikely to hold, although of course there could be other factors that operate largely independently of geography, or on scales smaller or much larger than the scale of individual precincts. For instance, there could be a sorting of political opinion along social networks that is contributing to polarization that is uncorrelated with the coarse-grained geographic distribution of political opinion. It should also be noted that this empirical analysis cannot give informa-

tion about how extreme individual opinions are, but rather reveals patterns regarding only an aggregate measure—the fraction of people voting for one candidate over the other—of the political opinion in each precinct. Overall though, this empirical analysis suggests that the theory presented should be taken as descriptive and that one should be skeptical of the casual conclusions that are drawn from it.

7.4 Conclusion

We have examined the spatial structure of U.S. voting patterns in the 2016 general presidential election, creating a predictive model based on demographics that suggests a lack of spatial causal influence of political opinions. Thus, while social influence may have a large impact on political opinion either within precincts or across precincts but across self-selected ties, it appears to operate in such a way that political opinion is not affected by the average opinion of neighboring precincts.

Nonetheless, there are substantial geospatial correlations in political opinion, both due to correlations in demographics and correlations from other factors. We examine the nature of the correlations both in demographics and from other factors besides the demographics we examined (the actual voting patterns being the sum of these two), demonstrating their multi-scale nature.

A key concept is the scale over which opinion is aggregated. When considering a federal governance structure with governments representing individuals at various levels, this scale becomes particularly relevant. As one example, the moderation of the U.S. Senate as compared with the House of Representatives can be explained by the fact that (even without gerrymandering) opinion aggregated at the state level will usually be more moderate than opinion aggregated at the congressional district level.

Our geospatial results are primarily descriptive and meant to spur further theoretical analysis. To illustrate one such possible direction, we consider the implications of this geospatial structure for multi-scale (e.g. local, state, federal) governance systems.

A fundamental principle in complex systems science is that the multi-scale complexity of the system should match that of the environment (section 1.1.5). However, the multi-scale structure of government is not at all aligned with the multi-scale structure we see in demographics and political opinion. Most power is concentrated at the federal level, resulting in a national government that can only represent part of the country at any one time (chapter 6). Yet lower level governance can also be misaligned with the geographic structure of political opinion—as can be seen, there is substantial variance in political opinion below the level of a state but above the level of individual towns and cities that cannot be well represented on many issues by state-wide policy.

Furthermore, as described in section 1.1.7, the states themselves do not form natural boundaries with respect to political opinion (with the exception, perhaps, of New Hampshire and Vermont): parts of one state are often more likely to resemble different parts of different states than they are to resemble each other. (For instance, political opinion in New York City resembles that in Chicago far more than it does upstate New York.) Thus, devolving power to the states, while providing some degree of self-determination and allowing for various policies to be tried out in parallel, may not be sufficient: to more fully capture the patterns in political opinion and reduce the extent to which the minorities must submit to majority rule, more decisions (which may currently be made at the state or federal level) may need to be made on a local level. Given the large number of localities, however, it may also be useful for locales to coordinate to some degree with similar locales, even across straight lines, thus allowing for larger-scale coordination that aligns more with the actual spatial structure of political opinion than the way in which larger-scale coordination traditionally occurs through state and federal governments.

There are many reasons why polarization is bad for democracy, perhaps the most fundamental of which is its hindrance of effective representation, especially in the absence of compulsory voting (see chapter 6). Some have argued that polarization occurs in multi-decade cycles, and that it will eventually resolve itself with a political realignment (as occurred with the New Deal coalition) [120]. However, the last period in which polarization was so tied

to geography ended with a civil war. Furthermore, due to the technology-fueled dissolution of local connections in favor of long-distance ones, the dynamics of political opinion today may be fundamentally different from those of the past. Further analysis is needed, but polarization is not guaranteed to dissipate without intervention, in which case we must directly combat it or find ways to govern in spite of it.

7.5 Appendix

7.5.1 Gaussian model

We are interested in capturing the general forces that shape political opinion, with an eye towards understanding the different roles that local and global interactions play. Most treatments of public opinion do not consider the effect of geographic location, or do so only to note that there are in fact geographic differences. They treat individuals either as independent—interacting with nobody—or subject to general cultural forces, interacting with everybody with equal probability.[9] But the effects of location go beyond simply differences, since people are more likely to interact with those who live nearby. We consider both idiosyncratic forces that independently shape opinion in each locale, as well as the forces of social context that arise from local social interactions.

Our model considers opinion as a function of space. Formally, people occupy a geographic position $x \in \mathbb{R}^d$ and have an opinion $m \in \mathbb{R}^n$. Since we are not so much interested in the behavior of individual people as we are in how opinion varies over space, we consider a continuous model of the average opinion of the people living in some radius r_0 around location x, which we denote by $m(x)$.[10] This radius r_0 must be chosen to be large enough such that $m(x)$ is smoothly varying and does not depend on the behavior of a few individuals.[11] This

[9] Both of these approximations are tightly related and are forms of the mean-field approximation.

[10] More precisely, $m(x)$ is the convolution of the opinion as a function of space with some kernel parameterized by the radius.

[11] The exact choice of r_0 over which opinions are averaged is arbitrary, in the sense that the results derived from the model should not depend on r_0. In order to maintain this condition, the parameters of the model, however, will be functions of r_0.

model is statistical, i.e. $m(x)$ is a random variable for each x, and in the absence of spatial interactions, we denote the expected value of the opinion at each location $\langle m(x)\rangle$ by $m_0(x)$. For simplicity, we assume that the opinion at each location has the same covariance $\sigma^2 I_n$, where I_n is the $n \times n$ identity matrix. (This covariance has nothing to do with the sample covariance of the opinions of the individuals living within r of x, but rather reflects our uncertainty in the value of the sample average.) Then, for any location x, the simplest probability distribution for the average opinion around that location is gaussian and given by $\mathbb{P}[m(x)] \propto \exp\left[-\frac{1}{2\sigma^2}m(x)^2 + h(x)m(x)\right]$, where $h(x) = m_0(x)/\sigma^2$. (The multiplication of two vectors in $\mathbb{R}^n$ should be interpreted as a dot-product.) Still under the assumption of no spatial interactions, the probability of finding a particular spatial configuration of opinion is then given by

$$\mathbb{P}[m] = \frac{1}{Z}e^{-f[m]} \tag{7.1}$$

where the normalization constant $Z = \sum_m e^{-f[m]}$ is known as the partition function[12] and f is defined by

$$f[m] = \int d^d x \left[\frac{1}{2\sigma^2}m(x)^2 - h(x)m(x)\right] \tag{7.2}$$

where the integral is taken over all $x \in \mathbb{R}^d$.[13] It can be verified that this choice of f yields[14] $\langle m(x)\rangle = \sigma^2 h(x) = m_0(x)$.

We then add in the simplest possible spatial interaction, an "energy" penalty of $\frac{K}{2}(\nabla m(x))^2$. This term makes configurations in which the opinion varies a lot over short distances less likely. Keeping the probability distribution in the form of eq. (7.1), the model is now de-

[12]Note that $\sum_m$ represents a sum over all possible configurations of $m(x)$, i.e. all possible functions of opinion over space. The probability that any given opinion configuration $m(x)$ occurs, $\mathbb{P}[m]$, is proportional to $e^{-f[m]}$, which means that we can loosely think of $f[m]$ as proportional to the energy of that configuration. Note that configurations with higher energy are less likely to occur. Up to an additive constant, $f[m]$ is also the negative log-probability of m. Energies/log-probabilities have the property that they are additive for noninteracting/independent events.

[13]Note: $\int d^d x$ is shorthand for $\int \Pi_{i=1}^d dx_i$ where x_i are the components of x.

[14]For this choice of f, we note that the values of m at each point x in space are independent random variables, and so we can look at the probability distribution for each point separately:

$$\mathbb{P}[m(x)] \propto e^{-\left[\frac{1}{2\sigma^2}m(x)^2 - h(x)m(x)\right]} = e^{-\frac{1}{2\sigma^2}\left(m(x)-\sigma^2 h(x)\right)^2} e^{\frac{1}{2}\sigma^2 h(x)^2}$$

Since this distribution is gaussian, the expected value of $m(x)$ can just be read off.

scribed by

$$f[m] = \int d^d x \left[\frac{1}{2\sigma^2} m(x)^2 + \frac{K}{2} (\nabla m(x))^2 - h(x) m(x) \right] \tag{7.3}$$

Performing the Fourier transforms $m(x) = \frac{1}{V} \sum_q \tilde{m}(q) e^{iqx}$ and $h(x) = \frac{1}{V} \sum_q \tilde{h}(q) e^{iqx}$ (with $V = \int d^d x$ being the length/area/volume of the system) yields

$$f[m] = \frac{1}{V} \sum_q \left[\frac{\sigma^{-2} + Kq^2}{2} |\tilde{m}(q)|^2 - \tilde{h}(q) \tilde{m}(-q) \right] \tag{7.4}$$

from which we find that the expected values of the opinion Fourier modes are given by[15]

$$\langle \tilde{m}(q) \rangle = \frac{\tilde{m}_0(q)}{1 + q^2 \xi^2} \tag{7.5}$$

where $\xi = \sqrt{\sigma^2 K}$ is known as the correlation length.[16]

7.5.2 Spatial self-sorting

Over the past several decades, people have been sorting themselves spatially, as described in Bill Bishop's *The Big Sort* [62]. Bishop argues that political polarization is a result of this sorting. Here, we show how polarization arises from sorting in our model.

Prior to self-sorting, we expect the opinions people would have in the non-interacting model (eq. (7.2)) to be uncorrelated for nearby points in space, and so $\tilde{m}_0(q)$ will be dominated by Fourier modes with $q\xi >> 1$. With spatial interactions, these Fourier modes are suppressed by a factor of $1 + q^2\xi^2$ (eq. (7.5)), and so, prior to self-sorting, natural differences in individual opinion will be largely suppressed. But after self-sorting, $\tilde{m}_0(q)$ will be dominated by Fourier modes such that $q\xi \sim 1$, which result in much less damping of individual opinions. This model is extremely simple, and does not include any non-linear effects

[15] The Fourier modes $\tilde{m}(q)$ are drawn from independent Gaussian distributions, allowing their expectation values to be calculated in the manner described in the previous footnote.

[16] ξ is called the correlation length because the correlation in opinion between two points that are far apart decays exponentially as $e^{-r/\xi}$ where r is the spatial distance between the two points.

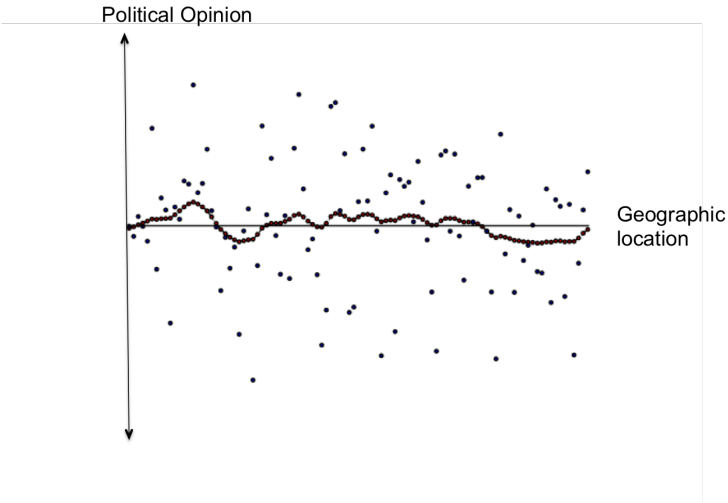


Figure 7-5: Before the Big Sort. We consider one spatial dimension (i.e. $d = 1$), represented by the horizontal direction of the figure, and one opinion dimension (i.e. $n = 1$), represented by the vertical direction. The expected value of the opinions $m_0(x)$ for the non-interacting model (eq. (7.2)) and the expected values of the opinions including the effects of interactions (eq. (7.3)) are shown in blue and red, respectively, for 101 points distributed evenly in the spatial dimension. These 101 values of $m_0(x)$ (blue) are independently drawn from a standard normal distribution, i.e. they have no spatial correlation. The lines connecting the red dots heuristically represent the interactions between people who are near each other. The horizontal black line is the spatial axis, corresponding to an opinion of 0.

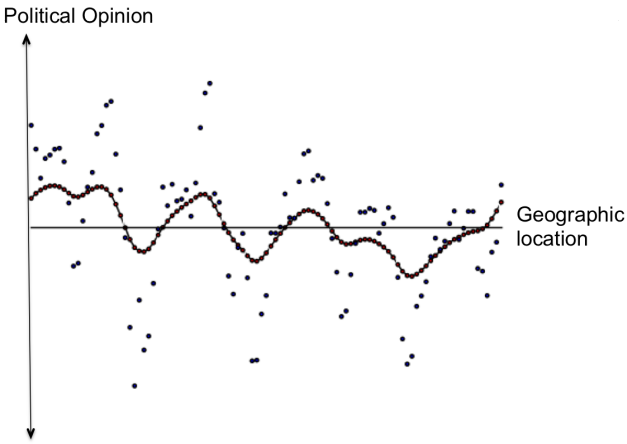


Figure 7-6: After some sorting. This is a figure of the same people as above after they have spatially sorted for 1000 time-steps. (In each time step, two individuals are randomly chosen, and they are swapped if and only if the swap results in both of them being closer in opinion to their new neighbors.) Note that the sorting does not change the values of $m_0(x)$—it just changes the spatial locations of these values.

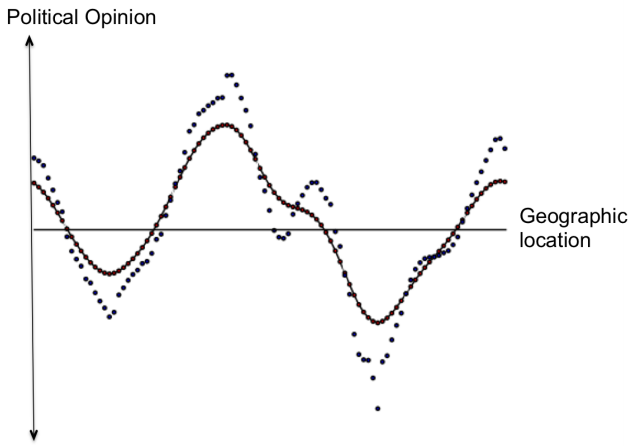


Figure 7-7: After more sorting. This is the opinion distribution after 100,000 time-steps.

such as group polarization,[17] but already we see that even in the simplest linear model, the self-sorting of like-minded individuals makes consensus much harder to reach.

Although spatial self-sorting of individuals leads to $\tilde{m}_0$ being dominated by lower frequency Fourier modes, other forces could also lead to the same mathematical effect. In other words, polarization will arise from the geographic clustering of similar innate forces, regardless of what the cause of this clustering is. Even if Bishop's argument is flawed, polarization due to geographic clustering could still arise through the following mechanisms.

- The presidential candidates have become more aligned with an already deepening urban-rural divide in the country.

- Bishop talks about a neighborhood by neighborhood sorting, as well as a large-scale urban-rural sorting (where liberals move to/remain in cities/surrounding suburbs and conservatives remain in rural areas or move to exurbs). However, many people do not even know their neighbors, so it is likely that only the larger-scale sorting is relevant to

[17]Group polarization refers to the observation that the average opinion on an issue of a set of people tends to become more extreme after they discuss the issue amongst themselves. If we consider the opinions after the individuals discuss amongst themselves to be a function of the opinions beforehand, then the linear part of this function can be described by this model (by a redefinition of the parameter $m_0(x)$), but any non-linearities will not be captured.

political polarization, consistent with a field-theory model with a minimum resolution on the scale of entire towns/cities.

- Exogenous forces (changing economy, changing cultural and religious forces, changing ideologies and emphases of both political parties, changing government policy) may be heightening an urban-rural divide by both increasing and making more salient the differences between rural and city living. (In the language of my model, the difference between urban and rural external fields is both increasing and becoming more aligned with the axis of polarization.)

All in all, rather than considering this sort of sorting as a general force, it would be more accurate and fruitful to look specifically at the urban-rural divide, which could be caused partly by sorting but also partly by an increasing alignment of politics, culture, and the economy along the divide.

7.5.3 External influences

We ask what is the effect of adding a term $-\int d^d x h_{ext}(x) m(x)$ to the Hamiltonian. The relevant quantity here is the susceptibility to external influence,

$$\chi(q) = \frac{\partial \langle m(q) \rangle}{\partial h_{ext}(q)} = \frac{\sigma^2}{1 + q^2 \xi^2} \tag{7.6}$$

The key thing to note here is that global influences have much more of an effect than local influences. Considering that local sources of influence such as local news and community groups are being replaced by global sources such as national media outlets and the internet, our society in general is now much more unstable.

7.5.4 Loss of locality

The parameter K is a measure of the strength of local interactions. A decrease in the strength of local interactions—a decrease in K—leads to larger differences and swings in

political opinion between localities, as seen in eqs. (7.5) and (7.6). A decrease in K can have a variety of causes, including

- A decrease in social capital (as described in Putnam's *Bowling Alone*) will reduce local social connections and influences.

- Self-selected social connections: various social or economic forces may result in connections that are more likely to be between like-minded people. These connections displace the random connections based on locality.

- Long-ranged connections can displace short-ranged connections. The internet and other technologies can reduce the degree to which we are connected with those who live near us.

7.5.5 Media effects

Thus far, media has been exogenous to the system. As in illustration of the generality of the field theory approach, we can consider what happens if we include media in the model. Denote the average bias of the news consumed at location x by $n(x)$. Then we can write the partition function as $Z = \sum_{m,n} e^{-f[m,n]}$, and we write down the simplest possible form for $f[m,n]$ that includes an interaction between media bias and public opinion:

$$f[m,n] = f[m] + \int d^d x \big[-m(x)n(x) + \alpha n(x)^2 \big] \tag{7.7}$$

Note that the units of $n(x)$ are chosen so that the coefficient of the $n(x)m(x)$ term is 1. As it turns out, it is possible to exactly sum over all configurations of $n(x)$ and write the partition function only as a sum over configurations of $m(x)$:

$$Z = \sum_{m,n} e^{-f[m,n]} = \sum_m \sum_n e^{-f[m,n]} = \sum_m e^{-\tilde{f}[m]} \tag{7.8}$$

with

$$\tilde{f}[m] = \int d^d x \left[(\frac{1}{2\sigma^2} - \frac{1}{2\alpha}) m(x)^2 + \frac{K}{2} (\nabla m(x))^2 - h(x) m(x) \right] \tag{7.9}$$

This $\tilde{f}[m]$ is in the same form as the $f[m]$ without considering the effect of media, except with σ^{-2} being replaced with $\sigma^{-2} - \alpha^{-1}$. Given that the $\alpha n(x)^2$ provides an energy penalty for news sources being too biased, we can identify α with some notion of journalistic integrity. Thus, that the effect of including media endogenously in the system is to increase the variance and variability of individual opinion, with lower levels of journalistic integrity leading to larger increases in the variability of opinion.

7.5.6 Beyond the Gaussian model

The model considered in section 7.5.1 has a free energy ($f[m]$) that is quadratic in the opinion-field $m(x)$, and so gives linear behavior. Although such a model can capture some of the basic properties of spatial opinion-dynamics, it fails to capture any non-linear effects. In this section, we add to our model terms beyond quadratic order and consider their effects. We also treat all heterogeneity as exogenous to the system; i.e., unlike in the previous section, we set $h(x) = 0$ for all x.

The most general choice of f that is spatially homogeneous and isotropic and contains only local terms consistent with rotational symmetry in opinion-space is

$$f[m] = \int d^d x [\frac{t}{2} m^2 + \frac{K}{2} (\nabla m)^2 + u_4 m^4 + u_6 m^6 + ... + v_4 m^2 (\nabla m)^2 + v_6 m^4 (\nabla m)^2 + ...] \tag{7.10}$$

We recall from section 7.5.1 that the parameters in a continuous model (t, K, u_4, u_6, etc. in the above model) depend on the radius r_0 over which opinions are averaged to give the continuous field $m(x)$. In the gaussian model, the parameters depended on r_0 in a very simple way, resulting in the global (large-scale) behavior of the system simply looking like a scaled-up version of the local (small-scale) behavior. But the introduction of non-linear terms, as in eq. (7.10) can cause a non-trivial flow of parameters as one looks at the system at larger and large scales, i.e. as one increases r_0. While there is generally very little we

can say about the local behavior of a system containing many higher-order terms, there are certain generalities we can make about the global behavior, because of the way in which such parameters scale.

For $d \geq 4$, it turns out that all terms except for t and K flow to zero as r_0 increases. In other words, although the non-quadratic terms can affect behavior at small scales, the large-scale behavior of the system will be well-described by a gaussian model. But for $d < 4$, no exact solution exists. However, there is a useful approximation scheme in which one can calculate the behavior of the system via a power-series expansion around $d = 4$, with $\epsilon = 4 - d$ as a small parameter. For $d = 3$ ($\epsilon = 1$), including even one term in the ϵ-expansion gives reasonably accurate results; for $d = 2$ ($\epsilon = 2$), such results provide qualitative insight.

Without getting into too many details of the expansion, the only non-linear term that we need to worry about for large-scale behavior is the u_4 term, since all-higher order terms flow to zero. Defining $u \equiv u_4$, we thus write

$$f[m] = \int d^d x \left[\frac{t}{2} m^2 + \frac{K}{2} (\nabla m)^2 + u m^4 \right] \tag{7.11}$$

The key insight that including the um^4 term gives us is that it is possible to look at political opinion with fine granularity, and see a highly polarized distribution with a hollowed-out political center (fig. 7-8), but when one decreases the granularity, the parameters transform and one is left with a more moderate political distribution (fig. 7-9). This can occur because as one zooms out and looks at larger and larger areas, the positive um^4 term results in an increase in the value of t that can result in a polarized distribution ($t < 0$) changing to an unpolarized one ($t > 0$).[18]

Essentially, in less than 4 spatial dimensions, we can get pockets of strong political opinion that arise from the interactions between individuals, while still at larger scales seeing moderate political districts. (Note that in this model, everyone has the same intrinsic political opinions: no individual is more likely to politically lean one way or another.) This

[18] The way in which the parameters change as a function of scale is captured by the renormalization group flow equations. To lowest order, the change in t as the scale (parameterized by l) is increased is given by $\frac{dt}{dl} = 2t + (4n + 8) \frac{K_d \Lambda^d}{t + K\Lambda^2} u$.

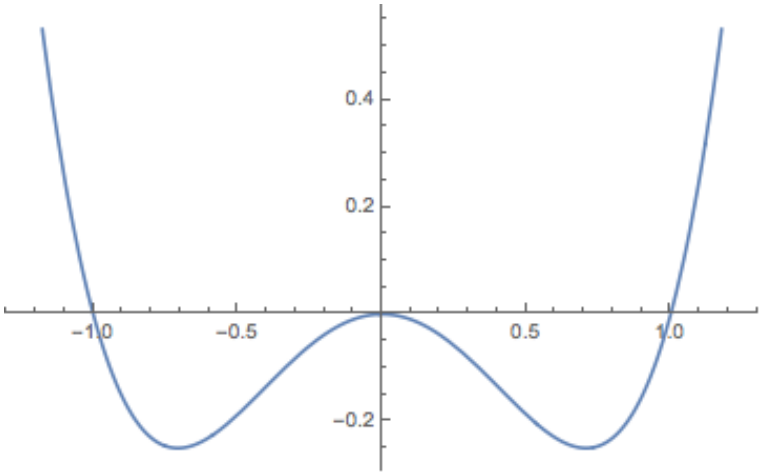


Figure 7-8: A "Mexican hat" potential ($t < 0$, $u > 0$). Note that extreme opinions in either direction are more likely than moderate ones.

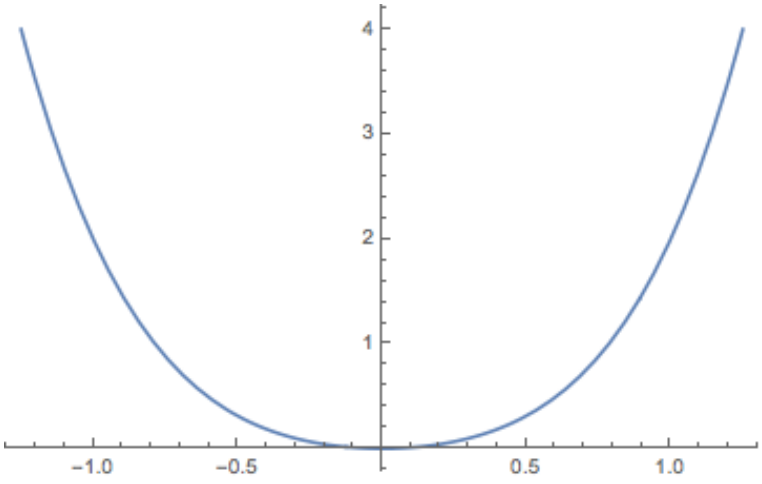


Figure 7-9: An unpolarized potential ($t > 0$, $u > 0$)

demonstrates a failure of mean-field theory, which predicts that society looks the same at all scales, and underscore the importance of considering how quantities of interest vary over space. Without going beyond mean-field theory, it is impossible to properly understand how the political center can be hollowed out.

For the regime in which t does eventually flow to positive values, we see that even though we started off with a non-linear model, we end up with a Gaussian model: once $t > 0$, u becomes less and less important compared to t as we further zoom out, and the results of section 7.5.1 will hold at sufficiently large scales. Whether or not a system that at fine scales appears polarized will end up in the Gaussian regime depends on the ratio $|t|/u$: for large enough values of $|t|$ compared to u, opinion will remain polarized even when aggregated at larger and larger spatial scales.

7.5.7 Self-selected long-ranged interactions

For long-ranged self-selected interactions, we consider terms up to the fourth order (this may require adding a vm^6 term to the local Hamiltonian for stability) that respect the symmetry $m \to -m$. We wish to define long-range interactions so that they do not affect the $q = 0$ mode (which is really a matter of preference, since we could just redefine t and u to compensate). We may also of course require that they be symmetric with respect to switching x and y, since the integration is symmetric. The lowest order term satisfying these conditions is $(m_x - m_y)^2$ term. We can then multiply this term by either $m_x m_y$ or $(m_x - m_y)^2$ yielding either $\int \frac{d^d x d^y}{2V} J_1 m_x m_y (m_x - m_y)^2$ or $\int \frac{d^d x d^y}{2V} J_2 (m_x - m_y)^4$. (Note that this formulation is equivalent to having two parameters to control $m_x^2 m_y^2$ and $m_x^3 m_y$ terms, which are the only two long-ranged fourth-order monomials.) We expect the effects of the J_2 term to be similar to those of the J term; thus we focus on the J_1 fourth-order term as a source of self-selected long-ranged interactions.

$$\beta H = \int d^d x [t m_x^2 + u m_x^4 + \int \frac{d^d y}{2V} (m_x - m_y)^2 (J + J_1 m_x m_y)] \tag{7.12}$$

where $u > J_1 \geq 0$, $J \geq 0$. We note that minimizing βH no longer necessarily means taking $m_x = \bar{m}$. Consider $m_x = \pm m$ with one half probability. Then

$$\frac{\beta H}{V} = tm^2 + um^4 + m^2(J - J_1 m^2) = (t + J)m^2 + (u - J_1)m^4 \tag{7.13}$$

Compare with choosing $m_x = m$, which yields $\beta H/V = tm^2 + um^4$. As long as $u > J_1$, this system is still stable overall, but for $t+J < 0$ and $t^2/u < (t+J)^2/(u-J_1)$, it becomes favorable for the population to split into two (and only two) components. (Note that $-J_1 t/u > 2J$ is a sufficient but not necessary condition.)

For a local Hamiltonian in which two locales are more likely to exhibit similar opinions when they are in close proximity, such spontaneous fragmentation in the absence of a symmetry-breaking external field cannot occur, since the lowest energy state of the Hamiltonian will be one where all locales have the same opinion. Thus, this spontaneous fragmentation cannot occur when interactions are predominantly local, which suggests a potential difference between the polarization of today and that of the past, the latter of which may have been driven by more "external" or top-down/centralized forces.

Chapter 8

Conclusion and Outlook

This thesis has introduced a number of complex systems principles, both directly and through examples of applications. It has also grounded the core ideas of complexity profiles and the multi-scale law of requisite variety in a novel mathematical formalism, modeled a strategy for stably eliminating pandemics with a combination of local control measures and travel restrictions, resolved a tension in the literature on surgical masks, and provided new concepts for thinking about democratic elections and political polarization. But what are the takeaway lessons?

In terms of general complex systems principles, a key lesson to draw attention to is the importance of matching the universality class of the model to the real-world system. A lot of complex systems science stems from the observation that although systems are often modeled (implicitly or explicitly) with assumptions of independence, they are not actually well-described by the mean-field-theory/central-limit-theorem universality class (section 1.2.1) that such models fall into. The response of many complex-systems-based analyses is to use some other tool (perhaps networks or agent-based models), which will yield a different universality class. However, such an approach ultimately suffers from the same problem as the independence-based analyses—it is grounded in a set of microscopic assumptions, which will inevitably get some things right and some things wrong, without sufficient consideration as to whether the system and the model have the same behaviors at larger scales.

While there can be value in such models, particularly in illustrating new types of large-scale behaviors that can emerge, it is difficult to distinguish between good and bad models based on microscopic assumptions alone. A model may seem like it matches the system reasonably well on a small-scale, yet yield drastically different behavior at a large scale (e.g. a model that left out electron-phonon interactions may seem relatively accurate since such interactions are so weak, yet could completely miss the onset of superconductivity). Conversely, a model could look completely different from the system at a small-scale, yet yield similar large-scale behavior (e.g., as explained by Fermi liquid theory, a free-electron model performs well for many metals despite its neglecting electron-electron interactions that are quite strong). Thus, an approach that is sensitive to the fact that a model cannot be evaluated based on its small-scale assumptions is needed.

As argued throughout this thesis, a universality-based approach, even if applied only conceptually, can help identify the important behaviors of systems and distinguish between good and bad models. This approach has a correspondence in rigorous analyses of the renormalization group in physics; however, a general version of the renormalization group has yet to be mathematically formalized for systems in general. Part of the difficulty is that even within physics, the formal application of the renormalization group is relatively narrow; for instance, a series of nested effective theories, each one more detailed than the last (e.g. a model of the physical properties of a material within which there are models of atoms, each of which could be described in terms of quantum fields), is conceptually related to but not mathematically formalized in terms of renormalization group flows. Thus, an important open question is: can the way in which multiple smaller-scale theories converge to the same larger-scale one be formalized in any general way?[1]

A key lesson from pandemic response is that models must be in service of our actions in the world, rather than our actions being slaves to our models. Implicit assumptions in models

[1]Of course, it is important to keep in mind that in physics, exact results are often obtained due to taking the limit as the ratio between two scales goes to infinity; for social systems, such precise results should not be expected, as the ratio between two scales will be finite. Thus, absolute statements about which parameters are relevant and irrelevant may become approximate when applied to social systems. There are of course numerous other challenges as well, including chaotic dynamics in which scales are amplified [50], and a general lack of perfect symmetry/isotropy/homogeneity.

can constrain the actions we consider; for example, a national model with only a single set of variables has already excluded internal travel restrictions from its state space. A good approach to modeling pandemics has to start with an evaluation of the space of possibilities (section 1.1), followed by an evaluation of their costs. For any potential pandemic, any region could start with the largest-scale question (fig. 3-1): "should we eliminate the disease?" For each possible answer to that question, various potential strategies could be considered, and within each strategy, various methods of implementation can be evaluated. Such an approach doesn't constrain the space of possibilities with narrow assumptions at a detailed level of modeling; rather, for each potential outcome, modeling is framed as a tool to estimate how such an outcome can be achieved. Within any large-scale framework, there is then lots of room for detailed research on how best to achieve certain aims (e.g. what is the least costly way to achieve a certain rate of exponential decline, what is the best way to prevent transmission of cases between regions, etc.). By estimating costs, certain branches of the decision tree can be ruled out. But by basing our models on our desired outcomes rather than vice versa, we avoid self-fulfilling prophecies in which model assumptions/predictions constrain the set of actions we consider.

For political systems, when considering proposed changes or evaluating interventions, their effect on electoral instability should be considered (chapter 6). That elections are unstable is often ignored: we take for granted that small changes in electorate behavior can have a big impact on the outcome. But it was not always this way: in 1950, for instance, the American Political Science Association lamented that the parties were too similar to one another and advocated that they polarize [198]. In fact, in order for elections to be stable, all candidates with a chance of winning the election must necessarily be relatively close together. (Of course, the positions of all candidates should change from election to election as the electorate changes.) In this way, small changes in the population can be sensitively and stably reflected in the outcome. More research is needed to precisely understand the effects of electoral reforms on stability: for instance, because party primaries can be a source of instability (section 6.3.4), a system such as instant-runoff/ranked-choice voting that eliminates

the need for primaries is promising. Even with a well designed electoral system, however, a diverse electorate can only be so well represented by a single instrument (chapter 7). More local governance may provide part of the solution, especially if localities can effectively coordinate to handle larger-scale challenges (section 1.1.7). Future work could include the design of stable, representative multi-scale (e.g. federal) electoral systems, as well as using the concepts developed in chapters 6 and 7 to comparatively study electoral systems across polities and time.

Election outcomes are influenced by electorate opinions, but, importantly, electorate opinions are also influenced by election outcomes and other large-scale forces such as the media and political campaigns/advertising. Our analysis has been limited to only part of this loop—the way in which electorate opinions influence outcomes. Through mechanisms by which election outcomes and other related large-scale forces influence electorate opinions, electoral instability could lead to further polarization, which would in turn lead to further instability in a self-reinforcing cycle. Reforms that reduce or eliminate this instability could help to break this cycle, but further research is needed to understand how elected officials and other "top-down" forces such as special interests interact with both each other and "bottom-up" social ties to shape the evolution of political opinion.

Ultimately, complex systems science is not based in any particular methodology but rather is an open-ended attempt to extend the scope of modeling beyond systems for which there are successful paradigms. As the field of physics has itself undergone multiple paradigm shifts and has been able to rigorously describe phenomena across many scales, many lessons from physics are applicable to studying complex systems more generally. A key pitfall must be avoided, however: rather than graft specific models from physics onto social systems (a line I perhaps cross in sections 7.2.1 and 7.5), we should instead use the conceptual framework that has made the study of physics so successful. In other words, we should imitate rather than copy. Much of complex systems science is making explicit what has previously been implicit, but I would be surprised if we ever find a complete framework for modeling complex systems—all explicit formulations are mere models: necessarily partial descriptions of an implicit whole.

Bibliography

[1] The subtle success of a complex mindset. *Nature Physics*, 14:1149, 2018.

[2] James Adams, Jay Dow, and Samuel Merrill. The political consequences of alienation-based and indifference-based voter abstention: Applications to presidential elections. *Political Behavior*, 28:65–86, 2006.

[3] Warish Ahmed, Nicola Angel, Janette Edson, Kyle Bibby, Aaron Bivins, Jake W. O'Brien, Phil M. Choi, Masaaki Kitajima, Stuart L. Simpson, Jiaying Li, Ben Tscharke, Rory Verhagen, Wendy J.M. Smith, Julian Zaugg, Leanne Dierens, Philip Hugenholtz, Kevin V. Thomas, and Jochen F. Mueller. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Science of The Total Environment*, 728:138764, 2020.

[4] Allison E. Aiello, Genevra F. Murray, Vanessa Perez, Rebecca M. Coulborn, Brian M. Davis, Monica Uddin, David K. Shay, Stephen H. Waterman, and Arnold S. Monto. Mask use, hand hygiene, and seasonal influenza-like illness among young adults: A randomized intervention trial. *The Journal of Infectious Diseases*, 201(4):491–498, 2010.

[5] Allison E. Aiello, Vanessa Perez, Rebecca M. Coulborn, Brian M. Davis, Monica Uddin, and Arnold S. Monto. Facemasks, hand hygiene, and influenza among young adults: A randomized intervention trial. *PLoS ONE*, 7(1):e29744, 2012.

[6] Marco Ajelli, Bruno Gonçalves, Duygu Balcan, Vittoria Colizza, Hao Hu, José J Ramasco, Stefano Merler, and Alessandro Vespignani. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 10(1):190, 2010.

[7] Fahad S. Al-Jasser, I. Kabbash, Mohammad A. Almazroa, and Z. Memish. Patterns of diseases and preventive measures among domestic hajjis from Central, Saudi Arabia. *Saudi medical journal*, 33 8:879–86, 2012.

[8] Julia E Aledort, Nicole Lurie, Jeffrey Wasserman, and Samuel A Bozzette. Non-pharmaceutical public health interventions for pandemic influenza: an evaluation of the evidence base. *BMC Public Health*, 7(1), 2007.

[9] Pietro Alessandrini and Michele U Fratianni. In the absence of fiscal union, the eurozone needs a more flexible monetary policy. *PSL Quarterly Review*, 68(275), 2015.

[10] Mohammad Alfelali, Elizabeth A. Haworth, Osamah Barasheed, Al-Mamoon Badahdah, Hamid Bokhary, Mohamed Tashani, Mohammad I. Azeem, Jen Kok, Janette Taylor, Elizabeth H. Barnes, Haitham El Bashir, Gulam Khandaker, Edward C. Holmes, Dominic E. Dwyer, Leon G. Heron, Godwin J. Wilson, Robert Booy, and Harunor Rashid and. Facemask against viral respiratory infections among hajj pilgrims: A challenging cluster-randomized trial. *PLOS ONE*, 15(10):e0240287, 2020.

[11] Sheikh Taslim Ali, Lin Wang, Eric H. Y. Lau, Xiao-Ke Xu, Zhanwei Du, Ye Wu, Gabriel M. Leung, and Benjamin J. Cowling. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science*, page eabc9004, 2020.

[12] Benjamin Allen, Blake Stacey, and Yaneer Bar-Yam. Multiscale information theory and the marginal utility of information. *Entropy*, 19(6):273, 2017.

[13] Linda J. S. Allen. An introduction to stochastic epidemic models. In Fred Brauer, Pauline van den Driessche, and Jianhong Wu, editors, *Mathematical Epidemiology*, pages 81–130. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[14] M. Alsved, A. Matamis, R. Bohlin, M. Richter, P-E. Bengtsson, C-J. Fraenkel, P. Medstrand, and J. Löndahl. Exhaled respiratory particles during singing and talking. *Aerosol Science and Technology*, pages 1–5, 2020.

[15] Benjamin M. Althouse, Edward A. Wenger, Joel C. Miller, Samuel V. Scarpino, Antoine Allard, Laurent Hébert-Dufresne, and Hao Hu. Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control. *PLOS Biology*, 18(11):e3000897, 2020.

[16] Philip W Anderson. More is different. *Science*, 177(4047):393–396, 1972.

[17] Viggo Andreasen. The final size of an epidemic and its relation to the basic reproduction number. *Bulletin of Mathematical Biology*, 73(10):2305–2321, 2011.

[18] Alex Arenas, Leon Danon, Albert Diaz-Guilera, Pablo M Gleiser, and Roger Guimera. Community analysis in social networks. *The European Physical Journal B*, 38(2):373–380, 2004.

[19] Kenneth J Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58:328–346, 1950.

[20] Kenneth J Arrow. *Social Choice and Individual Values*. Yale University Press, 2012.

[21] W Ross Ashby. *An Introduction to Cybernetics*. Chapman & Hall Ltd, 1961.

[22] W Ross Ashby. Requisite variety and its implications for the control of complex systems. In George J Klir, editor, *Facets of Systems Science*, pages 405–417. Springer Science & Business Media, 1991.

[23] Sangmin Aum, Sang Yoon (Tim) Lee, and Yongseok Shin. COVID-19 doesn't need lockdowns to destroy jobs: The effect of local outbreaks in Korea. Working Paper 27264, National Bureau of Economic Research, 2020.

[24] Robert Axelrod. *The Evolution of Cooperation*. Basic Books, 2006.

[25] Nihat Ay, Eckehard Olbrich, Nils Bertschinger, and Jürgen Jost. A unifying framework for complexity measures of finite systems. In *Proceedings of ECCS*, volume 6. Citeseer, 2006.

[26] Xuemei Bai, Ryan RJ McAllister, R Matthew Beaty, and Bruce Taylor. Urban policy and governance in a global environment: Complex systems, scale mismatches and public participation. *Current Opinion in Environmental Sustainability*, 2(3):129–135, 2010.

[27] Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Michele Tizzoni, Vittoria Colizza, and Alessandro Vespignani. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLOS One*, 6(1):e16591, 2011.

[28] Per Bak. *How Nature Works: The Science of Self-organized Criticality*. Springer Science & Business Media, 2013.

[29] Victor Balaban, William M. Stauffer, Adnan Hammad, Mohamud Afgarshe, Mohamed Abd-Alla, Qanta Ahmed, Ziad A. Memish, Janan Saba, Elizabeth Harton, Gabriel Palumbo, and Nina Marano. Protective practices and respiratory illness among US travelers to the 2009 hajj. *Journal of Travel Medicine*, 19(3):163–168, 2012.

[30] Fadoua Balabdaoui and Dirk Mohr. Age-stratified discrete compartment model of the COVID-19 epidemic with application to Switzerland. *Scientific Reports*, 10(1), 2020.

[31] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.

[32] Duygu Balcan, Bruno Gonçalves, Hao Hu, José J. Ramasco, Vittoria Colizza, and Alessandro Vespignani. Modeling the spatial spread of infectious diseases: The GLobal epidemic and mobility computational model. *Journal of Computational Science*, 1(3):132–145, 2010.

[33] Duygu Balcan and Alessandro Vespignani. Invasion threshold in structured populations with recurrent mobility patterns. *Journal of Theoretical Biology*, 293:87–100, 2012.

[34] Frank Ball, Tom Britton, Thomas House, Valerie Isham, Denis Mollison, Lorenzo Pellis, and Gianpaolo Scalia Tomba. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, 10:63–67, 2015.

[35] Frank Ball, Denis Mollison, and Gianpaolo Scalia-Tomba. Epidemics with two levels of mixing. *The Annals of Applied Probability*, 7(1):46–89, 1997.

[36] Philip Ball. *Why Society is a Complex Matter: Meeting Twenty-first Century Challenges with a New Kind of Science*. Springer Science & Business Media, 2012.

[37] José Balsa-Barreiro, Aymeric Vié, Alfredo J Morales, and Manuel Cebrián. Deglobalization in a hyper-connected world. *Palgrave Communications*, 6(1):1–4, 2020.

[38] Steven C Bankes. Tools and techniques for developing policies for complex and uncertain systems. *PNAS*, 99(suppl 3):7263–7266, 2002.

[39] Jeffrey S Banks and John Duggan. Probabilistic voting in the spatial model of elections: The theory of office-motivated candidates. In D. Austen-Smith and J. Duggan, editors, *Social Choice and Strategic Decisions*, pages 15–56. Springer, 2005.

[40] Yaneer Bar-Yam. *Dynamics of Complex Systems*. Addison-Wesley, 1997.

[41] Yaneer Bar-Yam. Complexity rising: From human beings to human civilization, a complexity profile. In Roberto Blancarte Pimentel, Robert Charles Elliot, Robert Holton, Pablo Lorenzano, and Herbert Arlt, editors, *Encyclopedia of Life Support Systems*, volume 1. EOLSS UNESCO Publishers, Oxford, UK, 2002.

[42] Yaneer Bar-Yam. Complexity of military conflict: Multiscale complex systems analysis of littoral warfare. *Report to Chief of Naval Operations Strategic Studies Group*, 2003.

[43] Yaneer Bar-Yam. *Making Things Work: Solving Complex Problems in a Complex World*. Knowledge Press, 2004.

[44] Yaneer Bar-Yam. A mathematical theory of strong emergence using multiscale variety. *Complexity*, 9(6):15–24, 2004.

[45] Yaneer Bar-Yam. Multiscale complexity/entropy. *Advances in Complex Systems*, 7(01):47–63, 2004.

[46] Yaneer Bar-Yam. Multiscale variety in complex systems. *Complexity*, 9(4):37–45, 2004.

[47] Yaneer Bar-Yam. Engineering complex systems: Multiscale analysis and evolutionary engineering. In Dan Braha, Ali A. Minai, and Yaneer Bar-Yam, editors, *Complex Engineered Systems*, pages 22–39. Springer, 2006.

[48] Yaneer Bar-Yam. Improving the effectiveness of health care and public health: A multiscale complex systems analysis. *American Journal of Public Health*, 96(3):459–466, 2006.

[49] Yaneer Bar-Yam. Complexity theory in applied policy worldwide. In Bernardo Alves Furtado, Patrícia AM Sakowski, and Marina H Tóvolli, editors, *Modeling Complex Systems for Public Policies*. Instituto de Pesquisa Econômica Aplicada, 2015.

[50] Yaneer Bar-Yam. From big data to important information. *Complexity*, 21(S2):73–98, 2016.

[51] Yavni Bar-Yam, Dion Harmon, and Yaneer Bar-Yam. Computationally tractable pairwise complexity profile. *Complexity*, 18(5):20–27, 2013.

[52] Albert-László Barabási. *Network Science*. Cambridge university press, 2016.

[53] Osamah Barasheed, Nedal Almasri, Al-Mamoon Badahdah, Leon Heron, Janette Taylor, Kenneth McPhee, Iman Ridda, Elizabeth Haworth, Dominic Dwyer, Harunor Rashid, and Robert Booy on behalf of the Hajj Research Team. Pilot randomised controlled trial to test effectiveness of facemasks in preventing influenza-like illness transmission among Australian hajj pilgrims in 2011. *Infectious Disorders - Drug Targets*, 14(2):110–116, 2014.

[54] Andrew Barbour and Denis Mollison. Epidemics and random graphs. In *Stochastic Processes in Epidemic Theory*, pages 86–89. Springer Berlin Heidelberg, 1990.

[55] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

[56] Stefano Benati and Giuseppe V Marzetti. Probabilistic spatial power indexes. *Social Choice and Welfare*, 40:391–410, 2013.

[57] Bernard Berelson, Paul Lazarsfeld, and William McPhee. *Voting: A Study of Opinion Formation in a Presidential Campaign*, chapter 6, pages 88–117. University of Chicago Press, 1954.

[58] Luis Bettencourt and Geoffrey West. A unified theory of urban living. *Nature*, 467(7318):912–913, 2010.

[59] Michel Bielecki, Roland Züst, Denise Siegrist, Daniele Meyerhofer, Giovanni Andrea Gerardo Crameri, Zeno Stanga, Andreas Stettbacher, Thomas Werner Buehrer, and Jeremy Werner Deuel. Social distancing alters the clinical course of COVID-19 in young adults: A comparative cohort study. *Clinical Infectious Diseases*, 2020.

[60] Md Arif Billah, Md Mamun Miah, and Md Nuruzzaman Khan. Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence. *PLOS ONE*, 15(11):e0242128, 2020.

[61] P-M Binder and Jorge A Plazas. Multiscale analysis of complex systems. *Physical Review E*, 63(6):065203, 2001.

[62] Bill Bishop. *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart*. Mariner Books, Boston, 2009.

[63] Duncan Black. On the rationale of group decision-making. *Journal of Political Economy*, 56:23–34, 1948.

[64] C. Makison Booth, M. Clayton, B. Crook, and J.M. Gawn. Effectiveness of surgical masks against influenza bioaerosols. *Journal of Hospital Infection*, 84(1):22–26, 2013.

[65] Christian Borghesi and Jean-Philippe Bouchaud. Spatial correlations in vote statistics: a diffusive field model for decision-making. *The European Physical Journal B*, 75:395–404, 2010.

[66] Christian Borghesi, Jean-Claude Raynal, and Jean-Philippe Bouchaud. Election turnout statistics in many countries: similarities, differences, and a diffusive field model for decision-making. *PloS One*, 7:e36289, 2012.

[67] Jean-Philippe Bouchaud. Crises and collective socio-economic phenomena: Simple models and challenges. *Journal of Statistical Physics*, 151(3-4):567–606, 2013.

[68] Dan Braha and Marcus AM de Aguiar. Voting contagion: Modeling and analysis of a century of US presidential elections. *PloS One*, 12:e0177970, 2017.

[69] Julii Suzanne Brainard, Natalia Jones, Iain Lake, Lee Hooper, and Paul Hunter. Facemasks and similar barriers to prevent respiratory illness such as COVID-19: A rapid systematic review. *medRxiv 2020.04.01.20049528*, 2020.

[70] Alan J Bray. Theory of phase-ordering kinetics. *Advances in Physics*, 51:481–587, 2002.

[71] Garry D Brewer. The challenges of interdisciplinarity. *Policy Sciences*, 32(4):327–337, 1999.

[72] Tom Britton, Frank Ball, and Pieter Trapman. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science*, 369(6505):846–849, 2020.

[73] James H Brown and Geoffrey B West, editors. *Scaling in Biology*. Oxford University Press, 2000.

[74] John S Brownstein, Cecily J Wolfe, and Kenneth D Mandl. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLOS Medicine*, 3(10):e401, 2006.

[75] Adam Brzezinski, Valentin Kecht, and David Van Dijcke. The cost of staying open: Voluntary social distancing and lockdowns in the US. Working Paper 910, University of Oxford, Department of Economics, 2020.

[76] Sergey V Buldyrev, Roni Parshani, Gerald Paul, H Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010.

[77] Henning Bundgaard, Johan Skov Bundgaard, Daniel Emil Tadeusz Raaschou-Pedersen, Christian von Buchwald, Tobias Todsen, Jakob Boesgaard Norsk, Mia M. Pries-Heje, Christoffer Rasmus Vissing, Pernille B. Nielsen, Ulrik C. Winsløw, Kamille Fogh, Rasmus Hasselbalch, Jonas H. Kristensen, Anna Ringgaard, Mikkel Porsborg Andersen, Nicole Bakkegård Goecke, Ramona Trebbien, Kerstin Skovgaard, Thomas Benfield, Henrik Ullum, Christian Torp-Pedersen, and Kasper Iversen. Effectiveness of adding a mask recommendation to other public health measures to prevent SARS-CoV-2 infection in danish mask wearers. *Annals of Internal Medicine*, 2020.

[78] David S Byrne. *Complexity Theory and the Social Sciences: An Introduction*. Psychology Press, 1998.

[79] Paul Cairney. Complexity theory in political science and public policy. *Political Studies Review*, 10(3):346–358, 2012.

[80] Randall L Calvert. Robustness of the multidimensional voting model: Candidate motivations, uncertainty, and convergence. *American Journal of Political Science*, 29:69–95, 1985.

[81] Laetitia Canini, Laurent Andréoletti, Pascal Ferrari, Romina D'Angelo, Thierry Blanchon, Magali Lemaitre, Laurent Filleul, Jean-Pierre Ferry, Michel Desmaizieres, Serge Smadja, Alain-Jacques Valleron, and Fabrice Carrat. Surgical mask to prevent influenza transmission in households: A cluster randomized trial. *PLoS ONE*, 5(11):e13998, 2010.

[82] Matthew B Canzoneri, Robert E Cumby, and Behzad T Diba. How do monetary and fiscal policy interact in the European monetary union? Technical report, National Bureau of Economic Research, 2005.

[83] Brian Castellani. Map of the complexity sciences. `https://www.art-sciencefactory.com/complexity-map_feb09.html`. Accessed 2020-05-15.

[84] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81:591–646, 2009.

[85] Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19) Frequently Asked Questions. https://www.cdc.gov/coronavirus/2019-ncov/faq.html, 2020. Accessed 2020-6-28.

[86] Jasper Fuk-Woo Chan, Shuofeng Yuan, Anna Jinxia Zhang, Vincent Kwok-Man Poon, Chris Chung-Sing Chan, Andrew Chak-Yiu Lee, Zhimeng Fan, Can Li, Ronghui Liang, Jianli Cao, Kaiming Tang, Cuiting Luo, Vincent Chi-Chung Cheng, Jian-Piao Cai, Hin

Chu, Kwok-Hung Chan, Kelvin Kai-Wang To, Siddharth Sridhar, and Kwok-Yung Yuen. Surgical mask partition reduces the risk of noncontact transmission in a golden syrian hamster model for coronavirus disease 2019 (COVID-19). *Clinical Infectious Diseases*, 2020.

[87] Arnab Chatterjee, Marija Mitrović, and Santo Fortunato. Universality in voting behavior: an empirical analysis. *Scientific Reports*, 3:1049, 2013.

[88] Vincent Chi-Chung Cheng, Shuk-Ching Wong, Vivien Wai-Man Chuang, Simon Yung-Chun So, Jonathan Hon-Kwan Chen, Siddharth Sridhar, Kelvin Kai-Wang To, Jasper Fuk-Woo Chan, Ivan Fan-Ngai Hung, Pak-Leung Ho, and Kwok-Yung Yuen. The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. *Journal of Infection*, 81(1):107–114, 2020.

[89] Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489):395–400, 2020.

[90] Sunhwa Choi and Moran Ki. Estimating the reproductive number and the outbreak size of COVID-19 in Korea. *Epidemiology and Health*, 42:e2020011, 2020.

[91] A. J. Choudhry, K. S. Al-Mudaimegh, A. M. Turkistani, and N. A. Al-Hamdan. Hajj-associated acute respiratory infection among hajjis from Riyadh. *East. Mediterr. Health J.*, 12(3-4):300–309, 2006.

[92] Rajiv Chowdhury, Kevin Heng, Md Shajedur Rahman Shawon, Gabriel Goh, Daisy Okonofua, Carolina Ochoa-Rosales, Valentina Gonzalez-Jaramillo, Abbas Bhuiya, Daniel Reidpath, Shamini Prathapan, Sara Shahzad, Christian L. Althaus, Nathalia Gonzalez-Jaramillo, and Oscar H. Franco. Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries. *European Journal of Epidemiology*, 35(5):389–399, 2020.

[93] Derek K Chu, Elie A Akl, Stephanie Duda, Karla Solo, Sally Yaacoub, Holger J Schünemann, Derek K Chu, Elie A Akl, Amena El-harakeh, Antonio Bognanni, Tamara Lotfi, Mark Loeb, Anisa Hajizadeh, Anna Bak, Ariel Izcovich, Carlos A Cuello-Garcia, Chen Chen, David J Harris, Ewa Borowiack, Fatimah Chamseddine, Finn Schünemann, Gian Paolo Morgano, Giovanna E U Muti Schünemann, Guang Chen, Hong Zhao, Ignacio Neumann, Jeffrey Chan, Joanne Khabsa, Layal Hneiny, Leila Harrison, Maureen Smith, Nesrine Rizk, Paolo Giorgi Rossi, Pierre AbiHanna, Rayane El-khoury, Rosa Stalteri, Tejan Baldeh, Thomas Piggott, Yuan Zhang, Zahra Saad, Assem Khamis, Marge Reinap, Stephanie Duda, Karla Solo, Sally Yaacoub, and Holger J Schünemann. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet*, 395(10242):1973–1987, 2020.

[94] Abrar A. Chughtai, Holly Seale, and C. Raina Macintyre. Effectiveness of cloth masks for protection against severe acute respiratory syndrome coronavirus 2. *Emerging Infectious Diseases*, 26(10), 2020.

[95] Pasquale Cirillo and Nassim Nicholas Taleb. Tail risk of contagious diseases. *Nature Physics*, 16(6):606–613, 2020.

[96] Vittoria Colizza, Romualdo Pastor-Satorras, and Alessandro Vespignani. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, 3(4):276–282, 2007.

[97] Vittoria Colizza and Alessandro Vespignani. Invasion threshold in heterogeneous metapopulation networks. *Phys. Rev. Lett.*, 99:148701, 2007.

[98] Vittoria Colizza and Alessandro Vespignani. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of Theoretical Biology*, 251(3):450–467, 2008.

[99] Peter Coughlin and Shmuel Nitzan. Electoral outcomes with probabilistic voting and Nash social welfare maxima. *Journal of Public Economics*, 15:113–121, 1981.

[100] Peter J Coughlin. *Probabilistic Voting Theory*. Cambridge University Press, 1992.

[101] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

[102] COVID19INDIA. `https://www.covid19india.org/`. Accessed: 2021-05-27.

[103] B. J. Cowling, Y. Zhou, D. K. M. Ip, G. M. Leung, and A. E. Aiello. Face masks to prevent transmission of influenza virus: a systematic review. *Epidemiology and Infection*, 138(4):449–456, 2010.

[104] Benjamin J. Cowling, Kwok-Hung Chan, Vicky J. Fang, Calvin K.Y. Cheng, Rita O.P. Fung, Winnie Wai, Joey Sin, Wing Hong Seto, Raymond Yung, Daniel W.S. Chu, Billy C.F. Chiu, Paco W.Y. Lee, Ming Chi Chiu, Hoi Che Lee, Timothy M. Uyeki, Peter M. Houck, J. S. Malik Peiris, and Gabriel M. Leung. Facemasks and hand hygiene to prevent influenza transmission in households. *Annals of Internal Medicine*, 151(7):437, 2009.

[105] Benjamin J. Cowling, Rita O. P. Fung, Calvin K. Y. Cheng, Vicky J. Fang, Kwok Hung Chan, Wing Hong Seto, Raymond Yung, Billy Chiu, Paco Lee, Timothy M. Uyeki, Peter M. Houck, J. S. Malik Peiris, and Gabriel M. Leung. Preliminary findings of a randomized trial of non-pharmaceutical interventions to prevent influenza transmission in households. *PLoS ONE*, 3(5):e2101, 2008.

[106] Mark C Cross and Pierre C Hohenberg. Pattern formation outside of equilibrium. *Reviews of Modern Physics*, 65(3):851, 1993.

[107] Horatiu Dan. The euro zone–between fiscal heterogeneity and monetary unity. *Transylvanian Review of Administrative Sciences*, 43 E:68–84, 2014.

[108] Alice Ware Davidson and Yaneer Bar-Yam. Environmental complexity: Information for human-environment well-being. In Ali A. Minai and Yaneer Bar-Yam, editors, *Unifying Themes in Complex Systems*, volume IIIB, pages 157–168. Springer, 2006.

[109] Anna Davies, Katy-Anne Thompson, Karthika Giri, George Kafatos, Jimmy Walker, and Allan Bennett. Testing the efficacy of homemade masks: Would they protect in an influenza pandemic? *Disaster Medicine and Public Health Preparedness*, 7(4):413–418, 2013.

[110] James B Davies, Rodrigo Lluberas, and Anthony F Shorrocks. Estimating the level and distribution of global wealth, 2000–2014. *Review of Income and Wealth*, 63(4):731–759, 2017.

[111] Nicholas G Davies, Adam J Kucharski, Rosalind M Eggo, Amy Gimma, W John Edmunds, Thibaut Jombart, Kathleen O'Reilly, Akira Endo, Joel Hellewell, Emily S Nightingale, Billy J Quilty, Christopher I Jarvis, Timothy W Russell, Petra Klepac, Nikos I Bosse, Sebastian Funk, Sam Abbott, Graham F Medley, Hamish Gibbs, Carl A B Pearson, Stefan Flasche, Mark Jit, Samuel Clifford, Kiesha Prem, Charlie Diamond, Jon Emery, Arminder K Deol, Simon R Procter, Kevin van Zandvoort, Yueqian Fiona Sun, James D Munday, Alicia Rosello, Megan Auzenbergs, Gwen Knight, Rein M G J Houben, and Yang Liu. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*, 5(7):e375–e385, 2020.

[112] Otto A Davis, Morris H DeGroot, and Melvin J Hinich. Social preference orderings and majority rule. *Econometrica: Journal of the Econometric Society*, 40:147–157, 1972.

[113] Otto A Davis, Melvin J Hinich, and Peter C Ordeshook. An expository development of a mathematical model of the electoral process. *American Political Science Review*, 64:426–448, 1970.

[114] Nicolas De Condorcet et al. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. de l'imprimerie royale, Paris, 1785.

[115] Pierre-Gilles de Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, 1979.

[116] J Bradford De Long, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann. Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Finance*, 45(2):379–395, 1990.

[117] Zakuan Zainy Deris, Habsah Hasan, Siti Amrah Sulaiman, Mohd Suhaimi Ab. Wahab, Nyi Nyi Naing, and Nor Hayati Othman. The prevalence of acute respiratory symptoms and role of protective measures among Malaysian hajj pilgrims. *Journal of Travel Medicine*, 17(2):82–88, 2010.

[118] O. Diekmann, J.A.P. Heesterbeek, and J.A.J. Metz. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4), 1990.

[119] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, 1990.

[120] Lawrence C. Dodd. Congress in a Downsian World: Polarization Cycles and Regime Change. *The Journal of Politics*, 77(2):311, 2015.

[121] Jean Dolbeault and Gabriel Turinici. Social heterogeneity and the COVID-19 lockdown in a multi-group SEIR model. *Computational and Mathematical Biophysics*, 9(1):14–21, 2021.

[122] Laura Di Domenico, Giulia Pullano, Chiara E. Sabbatini, Pierre-Yves Boëlle, and Vittoria Colizza. Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. *BMC Medicine*, 18(1), 2020.

[123] Anthony Downs. An economic theory of political action in a democracy. *Journal of Political Economy*, 65:135–150, 1957.

[124] James N. Druckman and Kjersten R. Nelson. Framing and Deliberation: How Citizens' Conversations Limit Elite Influence. *American Journal of Political Science*, 47(4):729, 2003.

[125] Francesco Drudi, Alain Durré, and Francesco Paolo Mongelli. The interplay of economic reforms and monetary policy: The case of the eurozone. *JCMS: Journal of Common Market Studies*, 50(6):881–898, 2012.

[126] Bruce Edmonds and Carlos Gershenson. Modelling complexity for policy: Opportunities and challenges. In Robert Geyer and Paul Cairney, editors, *Handbook on Complexity and Public Policy*. Edward Elgar Publishing, 2015.

[127] Manfred Eigen. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, 1971.

[128] Jon Elster and Aanund Hylland. *Foundations of Social Choice Theory*. Cambridge University Press, 1989.

[129] M. Emamian, Ali Hassani, and M. Fateh. Respiratory tract infections and its preventive measures among hajj pilgrims, 2010: A nested case control study. *International Journal of Preventive Medicine*, 4:1030 – 1035, 2013.

[130] Akira Endo, Sam Abbott, Adam J Kucharski, Sebastian Funk, et al. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research*, 5(67):67, 2020.

[131] James M Enelow and Melvin J Hinich. *The Spatial Theory of Voting: An Introduction*. Cambridge University Press, 1984.

[132] James M Enelow and Melvin J Hinich. A general probabilistic spatial theory of elections. *Public Choice*, 61:101–113, 1989.

[133] Joshua M Epstein, D Michael Goedecke, Feng Yu, Robert J Morris, Diane K Wagener, and Georgiy V Bobashev. Controlling pandemic flu: the value of international air travel restrictions. *PLOS One*, 2(5):e401, 2007.

[134] J Doyne Farmer, Mauro Gallegati, Cars Hommes, Alan Kirman, Paul Ormerod, Silvano Cincotti, Anxo Sanchez, and Dirk Helbing. A complex systems approach to constructing better models for managing financial markets and the economy. *The European Physical Journal Special Topics*, 214(1):295–324, 2012.

[135] Irwin Feller. New organizations, old cultures: Strategy and implementation of interdisciplinary programs. *Research Evaluation*, 11(2):109–116, 2002.

[136] Dan S Felsenthal and Moshé Machover. A priori voting power: what is it all about? *Political Studies Review*, 2:1–23, 2004.

[137] Neil M Ferguson, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial College COVID-19 Response Team*, 2020.

[138] Juan Fernández-Gracia, Krzysztof Suchecki, José J Ramasco, Maxi San Miguel, and Víctor M Eguíluz. Is the voter model a model for voters? *Physical Review Letters*, 112:158701, 2014.

[139] Santo Fortunato and Claudio Castellano. Scaling and universality in proportional elections. *Physical Review Letters*, 99:138701, 2007.

[140] Serge Galam. Sociophysics: A review of galam models. *International Journal of Modern Physics C*, 19:409–440, 2008.

[141] Monica Gandhi, Chris Beyrer, and Eric Goosby. Masks do more than protect others during COVID-19: Reducing the inoculum of SARS-CoV-2 to protect the wearer. *Journal of General Internal Medicine*, 2020.

[142] Alexander S Gard-Murray and Yaneer Bar-Yam. Complexity and the limits of revolution: What will happen to the Arab Spring? In Philip Vos Fellman, Yaneer Bar-Yam, and Ali A. Minai, editors, *Conflict and Complexity*, pages 281–292. Springer, 2015.

[143] J Gawn, M Clayton, C Makison, and B Crook. Evaluating the protection afforded by surgical masks against influenza bioaerosols: gross protection of surgical masks compared to filtering facepiece respirators. *Health Safety Exec*, 2008.

[144] Andrew Gelman, Jonathan N Katz, and Francis Tuerlinckx. The mathematics and statistics of voting power. *Statistical Science*, 17:420–435, 2002.

[145] Andrey Gerasimov, Georgy Lebedev, Mikhail Lebedev, and Irina Semenycheva. COVID-19 dynamics: A heterogeneous model. *Frontiers in Public Health*, 8, 2021.

[146] Elisabeth R Gerber and Jeffrey B Lewis. Beyond the median: Voter preferences, district heterogeneity, and political representation. *Journal of Political Economy*, 112:1364–1383, 2004.

[147] Timothy C. Germann, Kai Kadau, Ira M. Longini, and Catherine A. Macken. Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences*, 103(15):5935–5940, 2006.

[148] Robert Geyer and Paul Cairney. *Handbook on Complexity and Public Policy*. Edward Elgar Publishing, 2015.

[149] Speranta Gheorghiu-Svirschevski and Yaneer Bar-Yam. Multiscale analysis of information correlations in an infinite-range, ferromagnetic Ising system. *Physical Review E*, 70(6):066115, 2004.

[150] Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica: Journal of the Econometric Society*, 41:587–601, 1973.

[151] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[152] James Gleick. *Chaos: Making a New Science*. Open Road Media, 2011.

[153] Sholom Glouberman and Brenda Zimmerman. Complicated and complex systems: What would successful reform of Medicare look like? In *Romanow Papers: Changing Health Care in Canada*, volume 2, pages 21–53. University of Toronto Press, 2002.

[154] Nigel Goldenfeld and Carl Woese. Life is physics: Evolution as a collective phenomenon far from equilibrium. *Annual Review of Condensed Matter Physics*, 2(1):375–399, 2011.

[155] Brenda Goodman. The forgotten science behind face masks, 2020. Accessed 2021-02-04.

[156] Austan Goolsbee and Chad Syverson. Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline. Working Paper 27432, National Bureau of Economic Research, 2020.

[157] Wei Gou and Zhen Jin. How heterogeneous susceptibility and recovery rates affect the spread of epidemics on networks. *Infectious Disease Modelling*, 2(3):353–367, 2017.

[158] Andrzej Grabowski and RA Kosiński. Ising-based model of opinion formation in a complex network of interpersonal interactions. *Physica A: Statistical Mechanics and its Applications*, 361:651–664, 2006.

[159] Trisha Greenhalgh, Manuel B Schmid, Thomas Czypionka, Dirk Bassler, and Laurence Gruer. Face masks for the public during the COVID-19 crisis. *BMJ*, 369, 2020.

[160] Hermann Haken. *Information and Self-organization: A Macroscopic Approach to Complex Systems*. Springer Science & Business Media, 2006.

[161] Andrew G Haldane and Robert M May. Systemic risk in banking ecosystems. *Nature*, 469(7330):351–355, 2011.

[162] Andy Hall and Norman Clark. What do complex adaptive systems look like and what are the implications for innovation policy? *Journal of International Development*, 22(3):308–324, 2010.

[163] Michael Hammer and James Champy. *Reengineering the Corporation: A Manifesto for Business Revolution*. Zondervan, 2009.

[164] Dion Harmon, Marco Lagi, Marcus AM de Aguiar, David D Chinellato, Dan Braha, Irving R Epstein, and Yaneer Bar-Yam. Anticipating economic market crises using measures of collective panic. *PloS one*, 10:e0131871, 2015.

[165] Dion Harmon, Blake Stacey, Yavni Bar-Yam, and Yaneer Bar-Yam. Networks of economic market interdependence and systemic risk. *arXiv:1011.3707*, 2010.

[166] Joe Hasell. Which countries have protected both health and the economy in the pandemic? https://ourworldindata.org/covid-health-economy, 2020.

[167] Leila Hedayatifar, Rachel A. Rigg, Yaneer Bar-Yam, and Alfredo J. Morales. US social fragmentation at multiple scales. *Journal of The Royal Society Interface*, 16(159):20190509, 2019.

[168] Dirk Helbing. Globally networked risks and how to respond. *Nature*, 497(7447):51–59, 2013.

[169] Dirk Helbing, Dirk Brockmann, Thomas Chadefaux, Karsten Donnay, Ulf Blanke, Olivia Woolley-Meza, Mehdi Moussaid, Anders Johansson, Jens Krause, Sebastian Schutte, et al. Saving human lives: What complexity science and information systems can contribute. *Journal of Statistical Physics*, 158(3):735–781, 2015.

[170] Henry HQ Heng. The conflict between complex systems and reductionism. *JAMA*, 300(13):1580–1581, 2008.

[171] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.

[172] R.I. Hickson and M.G. Roberts. How population heterogeneity in susceptibility and infectivity influences epidemic dynamics. *Journal of Theoretical Biology*, 350:70–80, 2014.

[173] Melvin J Hinich. Some evidence on non-voting models in the spatial theory of electoral competition. *Public Choice*, 33:83–102, 1978.

[174] Melvin J Hinich, John O Ledyard, and Peter C Ordeshook. A theory of electoral equilibrium: A spatial analysis based on the theory of games. *The Journal of Politics*, 35:154–193, 1973.

[175] Melvin J Hinich, John O Ledyard, Peter C Ordeshook, et al. Nonvoting and the existence of equilibrium under majority rule. *Journal of Economic Theory*, 4:144–153, 1972.

[176] Dee Hock. *Birth of the Chaordic Age*. Berrett-Koehler Publishers, 1999.

[177] Harold Hotelling. Stability in competition. *The Economic Journal*, 39:41–57, 1929.

[178] Jeremy Howard, Austin Huang, Zhiyuan Li, Zeynep Tufekci, Vladimir Zdimal, Helene-Mari van der Westhuizen, Arne von Delft, Amy Price, Lex Fridman, Lei-Han Tang, Viola Tang, Gregory L. Watson, Christina E. Bax, Reshama Shaikh, Frederik Questier, Danny Hernandez, Larry F. Chu, Christina M. Ramirez, and Anne W. Rimoin. An evidence review of face masks against COVID-19. *Proceedings of the National Academy of Sciences*, 118(4):e2014564118, 2021.

[179] Robert Huckfeldt and John Sprague. Networks in Context: The Social Flow of Political Information. *American Political Science Review*, 84(4):1197, 1987.

[180] Alexandre Hyafil and David Moriña. Analysis of the impact of lockdown on the reproduction number of the SARS-cov-2 in spain. *Gaceta Sanitaria*, 35(5):453–458, 2021.

[181] Alvin J Ing, Christine Cocks, and Jeffery Peter Green. COVID-19: In the footsteps of Ernest Shackleton. *Thorax*, 75(8):693–694, 2020.

[182] Michael D Intriligator. A probabilistic model of social choice. *The Review of Economic Studies*, 40:553–560, 1973.

[183] Daniel J. Isenberg. Group Polarization: A Critical Review and Meta-analysis. *Journal of Personality and Social Psychology*, 50(6):1141, 1986.

[184] Istituto Superiore di Sanita. COVID-19 integrated surveillance: Key national data. https://www.epicentro.iss.it/en/coronavirus/sars-cov-2-integrated-surveillance-data. Accessed 2020-4-21.

[185] Shanto Iyengar and Kyu S. Hahn. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1):19, 2009.

[186] Jerry A Jacobs and Scott Frickel. Interdisciplinarity: A critical assessment. *Annual Review of Sociology*, 35:43–65, 2009.

[187] Irving L Janis. Groupthink. *IEEE Engineering Management Review*, 36(1):36, 2008.

[188] Tom Jefferson, Chris B Del Mar, Liz Dooley, Eliana Ferroni, Lubna A Al-Ansary, Ghada A Bawazeer, Mieke L Van Driel, Mark A Jones, Sarah Thorning, Elaine M Beller, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database of Systematic Reviews*, 2011(11):CD006207, 2011.

[189] Ron Johnston, David Manley, and Kelvyn Jones. Spatial Polarization of Presidential Voting in the United States, 1992-2012: The "Big Sort" Revisited. *Annals of the American Association of Geographers*, 106(5):1047, 2016.

[190] Rika Jolie, L Bäckström, and Chet Thomas. Health problems in veterinary students after visiting a commercial swine farm. *Canadian journal of veterinary research = Revue canadienne de recherche vétérinaire*, 62:44–8, 1998.

[191] Soren Jordan, Clayton M Webb, and B. Dan Wood. The president, polarization and the party platforms. *The Forum*, 12:169–189, 2014.

[192] Leo P Kadanoff. More is the same; phase transitions and mean field theories. *Journal of Statistical Physics*, 137(5-6):777, 2009.

[193] Mehran Kardar. *Statistical Physics of Fields*. Cambridge University Press, 2007.

[194] Stuart A Kauffman. *The Origins of Order: Self-organization and Selection in Evolution*. Oxford University Press, USA, 1993.

[195] Eben Kenah, Marc Lipsitch, and James M. Robins. Generation interval contraction and epidemic data analysis. *Mathematical Biosciences*, 213(1):71–79, 2008.

[196] Chris Kenyon. Widespread use of face masks in public may slow the spread of SARS-CoV-2: an ecological study. *medRxiv 2020.03.31.20048652*, 2020.

[197] Choon Ok Kim, Chung Mo Nam, Duk-Chul Lee, Joon Chang, and Ji Won Lee. Is abdominal obesity associated with the 2009 influenza A (H1N1) pandemic in Korean school-aged children? *Influenza and Other Respiratory Viruses*, 6(5):313–317, 2011.

[198] Evron M Kirkpatrick. "Toward a more responsible two-party system": Political science, policy science, or pseudo-science? *American Political Science Review*, 65(4):965–990, 1971.

[199] Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6:888–893, 2010.

[200] Nicole E Kogan, Leonardo Clemente, Parker Liautaud, Justin Kaashoek, Nicholas B Link, Andre T Nguyen, Fred S Lu, Peter Huybers, Bernd Resch, Clemens Havas, et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real-time. *arXiv:2007.00756*, 2020.

[201] Abhiteja Konda, Abhinav Prakash, Gregory A. Moss, Michael Schmoldt, Gregory D. Grant, and Supratik Guha. Aerosol filtration efficiency of common fabrics used in respiratory cloth masks. *ACS Nano*, 14(5):6339–6347, 2020.

[202] Jouni Korhonen and Thomas P Seager. Beyond eco-efficiency: A resilience perspective. *Business Strategy and the Environment*, 17(7):411–419, 2008.

[203] Moritz U. G. Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M. Pigott, null null, Louis du Plessis, Nuno R. Faria, Ruoran Li, William P. Hanage, John S. Brownstein, Maylis Layan, Alessandro Vespignani, Huaiyu Tian, Christopher Dye, Oliver G. Pybus, and Samuel V. Scarpino. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490):493–497, 2020.

[204] Arie W. Kruglanski and Ofra Mayseless. Motivational effects in the social comparison of opinions. *Journal of Personality and Social Psychology*, 53(5):834, 1987.

[205] Thomas S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 1962.

[206] Toshikazu Kuniya. Prediction of the epidemic peak of coronavirus disease in Japan, 2020. *Journal of Clinical Medicine*, 9(3):789, 2020.

[207] Yuliya N. Kyrychko, Konstantin B. Blyuss, and Igor Brovchenko. Mathematical modelling of the dynamics and containment of COVID-19 in Ukraine. *Scientific Reports*, 10(1), 2020.

[208] Elaine L. Larson, Yu-Hui Ferng, Jennifer Wong-McLoughlin, Shuang Wang, Michael Haber, and Stephen S. Morse. Impact of non-pharmaceutical interventions on URIs

and influenza in crowded, urban households. *Public Health Reports*, 125(2):178–191, 2010.

[209] Joseph T.F. Lau, Mason Lau, Jean H. Kim, Eric Wong, Hi-Yi Tsui, Thomas Tsang, and Tze Wai Wong. Probable secondary infections in households of SARS patients in Hong Kong. *Emerging Infectious Diseases*, 10(2):236–243, 2004.

[210] Joseph T.F. Lau, Hiyi Tsui, Mason Lau, and Xilin Yang. SARS transmission, risk factors, and prevention in Hong Kong. *Emerging Infectious Diseases*, 10(4):587, 2004.

[211] John O Ledyard. The pure theory of large two-candidate elections. *Public Choice*, 44:7–41, 1984.

[212] Christopher T. Leffler, Edsel Ing, Joseph D. Lykins, Matthew C. Hogan, Craig A. McKeown, and Andrzej Grzybowski. Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks. *The American Journal of Tropical Medicine and Hygiene*, 103(6):2400 – 2411, 2020.

[213] Sharachchandra Lélé and Richard B Norgaard. Practicing interdisciplinarity. *BioScience*, 55(11):967–975, 2005.

[214] Robert J Lempert. A new decision sciences for complex systems. *PNAS*, 99(suppl 3):7309–7313, 2002.

[215] Nancy H. L. Leung, Daniel K. W. Chu, Eunice Y. C. Shiu, Kwok-Hung Chan, James J. McDevitt, Benien J. P. Hau, Hui-Ling Yen, Yuguo Li, Dennis K. M. Ip, J. S. Malik Peiris, Wing-Hong Seto, Gabriel M. Leung, Donald K. Milton, and Benjamin J. Cowling. Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nature Medicine*, 26(5):676–680, 2020.

[216] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al. Early transmission dynamics in Wuhan, China of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382:1199–1207, 2020.

[217] Bernard Lietaer, Robert E Ulanowicz, Sally J Goerner, and Nadia McLaren. Is our monetary structure a systemic cause for financial instability? Evidence and remedies from nature. *Journal of Futures Studies*, 14(3):89–108, 2010.

[218] May Lim, Richard Metzler, and Yaneer Bar-Yam. Global pattern formation and ethnic/cultural violence. *Science*, 317(5844):1540–1544, 2007.

[219] Tse-Min Lin, James M Enelow, and Han Dorussen. Equilibrium in multicandidate probabilistic spatial voting. *Public Choice*, 98:59–82, 1999.

[220] Assar Lindbeck and Jörgen W Weibull. A model of political equilibrium in a representative democracy. *Journal of Public Economics*, 51:195–209, 1993.

[221] William G Lindsley, William P King, Robert E Thewlis, Jeffrey S Reynolds, Kedar Panday, Gang Cao, and Jonathan V Szalajda. Dispersion and exposure to a cough-generated aerosol in a simulated medical examination room. *Journal of Occupational and Environmental Hygiene*, 9(12):681–690, 2012.

[222] Lewis A Lipsitz. Understanding health care as a complex system: The foundation for unintended consequences. *JAMA*, 308(3):243–244, 2012.

[223] Christian List. Social choice theory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2013.

[224] James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.

[225] Milton Lodge and Charles S. Taber. *The Rationalizing Voter*. Cambridge University Press, 2013.

[226] Donald Ludwig, Brian Walker, and Crawford S Holling. Sustainability, stability, and resilience. *Conservation Ecology*, 1(1), 1997.

[227] Greg Lukianoff and Jonathan Haidt. The coddling of the American mind. *The Atlantic*, 316(2):42–52, 2015.

[228] Issa Luna-Pla and José R Nicolás-Carlock. Corruption and complexity: A scientific framework for the analysis of corruption networks. *Applied Network Science*, 5(1):1–18, 2020.

[229] Thomas Lux and Michele Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500, 1999.

[230] Wei Lyu and George L. Wehby. Community use of face masks and COVID-19: Evidence from a natural experiment of state mandates in the US. *Health Affairs*, 39(8):1419–1425, 2020.

[231] Junling Ma and David J. D. Earn. Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of Mathematical Biology*, 68(3):679–702, 2006.

[232] C. Raina MacIntyre, Simon Cauchemez, Dominic E. Dwyer, Holly Seale, Pamela Cheung, Gary Browne, Michael Fasher, James Wood, Zhanhai Gao, Robert Booy, and Neil Ferguson. Face mask use and control of respiratory virus transmission in households. *Emerging Infectious Diseases*, 15(2):233–241, 2009.

[233] Chandini Raina MacIntyre, Yi Zhang, Abrar Ahmad Chughtai, Holly Seale, Daitao Zhang, Yanhui Chu, Haiyan Zhang, Bayzidur Rahman, and Quanyi Wang. Cluster randomised controlled trial to examine medical mask use as source control for people with respiratory illness. *BMJ Open*, 6(12):e012330, 2016.

[234] Alexander V. Maltsev and Michael D. Stern. Social heterogeneity drives complex patterns of the COVID-19 pandemic: Insights from a novel stochastic heterogeneous epidemic model (SHEM). *Frontiers in Physics*, 8, 2021.

[235] Apoorva Mandavilli. It's not whether you were exposed to the virus. It's how much. *The New York Times*, May 2020.

[236] Benoit B Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Co., 1982.

[237] Rosario N Mantegna and H Eugene Stanley. *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, 1999.

[238] Spiro Maroulis, Roger Guimera, Hisham Petry, MJ Stringer, LM Gomez, LAN Amaral, and Uri Wilensky. Complex systems view of educational policy research. *Science*, 330(6000):38–39, 2010.

[239] Robert M May, Simon A Levin, and George Sugihara. Complex systems: Ecology for bankers. *Nature*, 451(7181):893–894, 2008.

[240] Michael P. McDonald. "National general election VEP turnout rates, 1789-present." *United States Election Project*, electproject.org/national-1789-present. Accessed 2019-7-25.

[241] Mark D McDonnell and Lawrence M Ward. The benefits of noise in neural systems: Bridging theory and experiment. *Nature Reviews Neuroscience*, 12(7):415, 2011.

[242] Richard D McKelvey and John W Patty. A theory of voting in large elections. *Games and Economic Behavior*, 57:155–180, 2006.

[243] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27:415–444, 2001.

[244] Matthew J. Memoli, Lindsay Czajkowski, Susan Reed, Rani Athota, Tyler Bristol, Kathleen Proudfoot, Sarah Fargis, Matthew Stein, Rebecca L. Dunfee, Pamela A. Shaw, Richard T. Davey, and Jeffery K. Taubenberger. Validation of the Wild-type Influenza A Human Challenge Model H1N1pdMIST: An A(H1N1)pdm09 Dose-Finding Investigational New Drug Study. *Clinical Infectious Diseases*, 60(5):693–702, 2014.

[245] Richard Metzler and Yaneer Bar-Yam. Multiscale complexity of correlated gaussians. *Physical Review E*, 71(4):046114, 2005.

[246] Robert A Meyers. *Encyclopedia of Complexity and Systems Science*. Springer, 2009.

[247] Quentin Michard and Jean-Philippe Bouchaud. Theory of collective opinion shifts: from smooth trends to abrupt swings. *The European Physical Journal B-Condensed Matter and Complex Systems*, 47:151–159, 2005.

[248] Joel C. Miller. A note on the derivation of epidemic final sizes. *Bulletin of Mathematical Biology*, 74(9):2125–2141, 2012.

[249] John H Miller and Scott E Page. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press, 2009.

[250] Donald K. Milton, M. Patricia Fabian, Benjamin J. Cowling, Michael L. Grantham, and James J. McDevitt. Influenza virus aerosols in human exhaled breath: Particle size, culturability, and effect of surgical masks. *PLoS Pathogens*, 9(3):e1003205, 2013.

[251] Melanie Mitchell. *Complexity: A Guided Tour*. Oxford University Press, 2009.

[252] Thierry Mora and William Bialek. Are biological systems poised at criticality? *Journal of Statistical Physics*, 144(2):268–302, 2011.

[253] Stefan Napel and Mika Widgrén. Power measurement as sensitivity analysis: a unified approach. *Journal of Theoretical Politics*, 16:517–538, 2004.

[254] Moti Nissani. Ten cheers for interdisciplinarity: The case for interdisciplinary knowledge and research. *The Social Science Journal*, 34(2):201–216, 1997.

[255] Joseph Norman and Yaneer Bar-Yam. Special operations forces as a global immune system. In Georgi Yordanov Georgiev, John M. Smart, Claudio L. Flores Martinez, and Michael E. Price, editors, *Evolution, Development and Complexity*, pages 367–379. Springer International Publishing, 2019.

[256] Torsten Olszak, Dingding An, Sebastian Zeissig, Miguel Pinilla Vera, Julia Richter, Andre Franke, Jonathan N Glickman, Reiner Siebert, Rebecca M Baron, Dennis L Kasper, et al. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science*, 336(6080):489–493, 2012.

[257] Elinor Ostrom. Beyond markets and states: Polycentric governance of complex economic systems. *American Economic Review*, 100(3):641–72, 2010.

[258] Guillermo Owen and Lloyd S Shapley. Optimal location of candidates in ideological space. *International Journal of Game Theory*, 18:339–356, 1989.

[259] Wei Pan, Yaniv Altshuler, and Alex Pentland. Decoding social influence and the wisdom of the crowd in financial trading network. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, pages 203–209. IEEE, 2012.

[260] Varuneswara Panyam, Hao Huang, Bogdan Pinte, Katherine Davis, and Astrid Layton. Bio-inspired design for robust power networks. In *2019 IEEE Texas Power and Energy Conference (TPEC)*, pages 1–6. IEEE, 2019.

[261] Sang Woo Park, Kaiyuan Sun, David Champredon, Michael Li, Benjamin M Bolker, David J. D. Earn, Joshua S Weitz, Bryan T Grenfell, and Jonathan Dushoff. Cohort-based approach to understanding the roles of generation and serial intervals in shaping epidemiological dynamics. *medRxiv 2020.06.04.20122713*, 2020.

[262] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925–979, 2015.

[263] Lionel S Penrose. The elementary statistics of majority voting. *Journal of the Royal Statistical Society*, 109:53–57, 1946.

[264] Pew Research Center. Political polarization in the American public. http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public, 2014.

[265] Pew Research Center. Electorally competitive counties have grown scarcer in recent decades. pewresearch.org/fact-tank/2016/06/30/electorally-competitive-counties-have-grown-scarcer-in-recent-decades, 2016.

[266] Pew Research Center. Political Polarization, 1994-2017. http://www.people-press.org/interactives/political-polarization-1994-2017/, 2017.

[267] Massimo Pizzol, Marco Scotti, and Marianne Thomsen. Network analysis as a tool for assessing environmental sustainability: Applying the ecosystem perspective to a Danish Water Management System. *Journal of Environmental Management*, 118:21–31, 2013.

[268] Dennis L Plane and Joseph Gershtenson. Candidates' ideological locations, abstention, and turnout in us midterm senate elections. *Political Behavior*, 26:69–93, 2004.

[269] Charles R Plott. A notion of equilibrium and its possibility under majority rule. *The American Economic Review*, 57:787–806, 1967.

[270] Chiara Poletto, Marcelo FC Gomes, Ana Pastore y Piontti, Luca Rossi, Livio Bioglio, Dennis L Chao, Ira M Longini, M Elizabeth Halloran, Vittoria Colizza, and Alessandro Vespignani. Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. *Euro Surveillance*, 19(42):20936, 2014.

[271] Keith T Poole and Howard Rosenthal. A spatial model for legislative roll call analysis. *American Journal of Political Science*, 29:357–384, 1985.

[272] Markus Prior. Media and Political Polarization. *Annual Review of Political Science*, 15:101, 2013.

[273] Robert D. Putnam. *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, New York, 2000.

[274] Jacob Ratkiewicz, Santo Fortunato, Alessandro Flammini, Filippo Menczer, and Alessandro Vespignani. Characterizing and modeling the dynamics of online popularity. *Physical Review Letters*, 105:158701, 2010.

[275] Erik M Rauch and Yaneer Bar-Yam. Long-range interactions and evolutionary stability in a predator-prey system. *Physical Review E*, 73(2):020903, 2006.

[276] Jonathan M. Read, Jessica R.E. Bridgen, Derek A.T. Cummings, Antonia Ho, and Chris P. Jewell. Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions. *medRxiv 2020.01.23.20018549*, 2020.

[277] Diana Rhoten and Andrew Parker. Risks and rewards of an interdisciplinary research path. *Science*, 306(5704):2046–2046, 2004.

[278] Julien Riou and Christian L Althaus. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance*, 25(4):2000058, 2020.

[279] Mick Roberts, Viggo Andreasen, Alun Lloyd, and Lorenzo Pellis. Nine challenges for deterministic epidemic models. *Epidemics*, 10:49–53, 2015.

[280] Kat Rock, Sam Brand, Jo Moir, and Matt J Keeling. Dynamics of infectious diseases. *Reports on Progress in Physics*, 77(2):026602, 2014.

[281] Alex Rutherford, Dion Harmon, Justin Werfel, Alexander S Gard-Murray, Shlomiya Bar-Yam, Andreas Gros, Ramon Xulvi-Brunet, and Yaneer Bar-Yam. Good fences: The importance of setting boundaries for peaceful coexistence. *PLOS One*, 9(5):e95660, 2014.

[282] Harry Rutter, Natalie Savona, Ketevan Glonti, Jo Bibby, Steven Cummins, Diane T Finegood, Felix Greaves, Laura Harper, Penelope Hawe, Laurence Moore, et al. The need for a complex systems model of evidence for public health. *The Lancet*, 390(10112):2602–2604, 2017.

[283] Anca Rădulescu, Cassandra Williams, and Kieran Cavanagh. Management strategies in a SEIR-type model of COVID 19 community spread. *Scientific Reports*, 10(1), 2020.

[284] Mark Allen Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.

[285] Robert Savit, Radu Manuca, and Rick Riolo. Adaptive competition, market efficiency, and phase transitions. *Physical Review Letters*, 82:2203–2206, 1999.

[286] Hiroki Sayama. *Introduction to the Modeling and Analysis of Complex Systems*. Open SUNY Textbooks, 2015.

[287] Antonio Scala, Andrea Flori, Alessandro Spelta, Emanuele Brugnoli, Matteo Cinelli, Walter Quattrociocchi, and Fabio Pammolli. Time, space and social interactions: Exit mechanisms for the COVID-19 epidemics. *Scientific Reports*, 10:13764, 2020.

[288] Marten Scheffer, Jordi Bascompte, William A Brock, Victor Brovkin, Stephen R Carpenter, Vasilis Dakos, Hermann Held, Egbert H Van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461:53–59, 2009.

[289] Marten Scheffer, Stephen R Carpenter, Timothy M Lenton, Jordi Bascompte, William Brock, Vasilis Dakos, Johan Van de Koppel, Ingrid A Van de Leemput, Simon A Levin, Egbert H Van Nes, et al. Anticipating critical transitions. *Science*, 338(6105):344–348, 2012.

[290] Thomas C Schelling. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2):143–186, 1971.

[291] Thomas C Schelling. *Micromotives and Macrobehavior*. WW Norton & Company, 2006.

[292] Frank Schweitzer, Giorgio Fagiolo, Didier Sornette, Fernando Vega-Redondo, Alessandro Vespignani, and Douglas R White. Economic networks: The new challenges. *Science*, 325(5939):422–425, 2009.

[293] Mark D Seery. Resilience: A silver lining to experiencing adverse life events? *Current Directions in Psychological Science*, 20(6):390–394, 2011.

[294] Willi Semmler and Wenlang Zhang. Monetary and fiscal policy interactions in the euro area. *Empirica*, 31(2-3):205–227, 2004.

[295] Lloyd S Shapley and Martin Shubik. A method for evaluating the distribution of power in a committee system. *The American Political Science Review*, 48:787–792, 1954.

[296] K. Shin, H. Wakabayashi, C. Sugita, Hiroki Yoshida, Keizo Sato, Tohru Sonoda, K. Yamauchi, F. Abe, and Masahiko Kurokawa. Effects of orally administered lactoferrin and lactoperoxidase on symptoms of the common cold. *International Journal of Health Sciences*, 12:44 – 50, 2018.

[297] Alexander F. Siegenfeld and Yaneer Bar-Yam. The impact of travel and timing in eliminating COVID-19. *Communications Physics*, 3:204, 2020.

[298] Alexander F. Siegenfeld and Yaneer Bar-Yam. An introduction to complex systems science and its applications. *Complexity*, 2020:6105872, 2020.

[299] Alexander F. Siegenfeld and Yaneer Bar-Yam. Negative representation and instability in democratic elections. *Nature Physics*, 16:186–190, 2020.

[300] Alexander F. Siegenfeld, Nassim Nicholas Taleb, and Yaneer Bar-Yam. What models can and cannot tell us about COVID-19. *Proceedings of the National Academy of Sciences*, 117(28):16092–16095, 2020.

[301] James M. Simmerman, Piyarat Suntarattiwong, Jens Levy, Richard G. Jarman, Suchada Kaewchana, Robert V. Gibbons, Ben J. Cowling, Wiwan Sanasuttipun, Susan A. Maloney, Timothy M. Uyeki, Laurie Kamimoto, and Tawee Chotipitayasunondh. Findings from a household randomized controlled trial of hand washing and face masks to reduce influenza transmission in Bangkok, Thailand. *Influenza and Other Respiratory Viruses*, 5(4):256–267, 2011.

[302] Herbert A Simon. *The Sciences of the Artificial*. MIT Press, 2019.

[303] Elliott Sober and David Sloan Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Harvard University Press, 1999.

[304] Didier Sornette. Physics and financial economics (1776–2014): Puzzles, Ising and agent-based models. *Reports on progress in physics*, 77:062001, 2014.

[305] Didier Sornette. *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton University Press, 2017.

[306] Priscilla L Southwell. The politics of alienation: Nonvoting and support for third-party candidates among 18–30-year-olds. *The Social Science Journal*, 40:99–107, 2003.

[307] Ralph D Stacey. *Complexity and Creativity in Organizations.* Berrett-Koehler Publishers, 1996.

[308] Dietrich Stauffer. Social applications of two-dimensional Ising models. *American Journal of Physics*, 76:470–473, 2008.

[309] John Sterman. *Business Dynamics*. Irwin/McGraw-Hill, 2010.

[310] Steven H Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. CRC Press, 2018.

[311] David Peter Stroh. *Systems Thinking for Social Change: A Practical Guide to Solving Complex Problems, Avoiding Unintended Consequences, and Achieving Lasting Results*. Chelsea Green Publishing, 2015.

[312] Laura F Su, Brian A Kidd, Arnold Han, Jonathan J Kotzin, and Mark M Davis. Virus-specific CD4+ memory-phenotype T cells are abundant in unexposed adults. *Immunity*, 38(2):373–383, 2013.

[313] Thorsten Suess, Cornelius Remschmidt, Susanne B Schink, Brunhilde Schweiger, Andreas Nitsche, Kati Schroeder, Joerg Doellinger, Jeanette Milde, Walter Haas, Irina

Koehler, Gérard Krause, and Udo Buchholz. The role of facemasks and hand hygiene in the prevention of influenza transmission in households: results from a cluster randomised trial; Berlin, Germany, 2009-2011. *BMC Infectious Diseases*, 12(1), 2012.

[314] Anthony D. Sung, Julia A.M. Sung, Kelly Corbet, Gloria Broadwater, Vera Hars, Ashley Zanter, Nelson J. Chao, and Mitchell E. Horwitz. Surgical mask usage reduces the incidence of parainfluenza virus 3 in recipients of stem cell transplantation. *Blood*, 120(21):462–462, 2012.

[315] Cass R. Sunstein. *Going to Extremes: How Like Minds Unite and Divide*. Oxford University Press, 2009.

[316] Jesse Sussell. New support for the big sort hypothesis: An assessment of partisan geographic sorting in California, 1992-2010. *PS: Political Science & Politics*, 46(4):768, 2013.

[317] Gin Nam Sze To and Christopher Yu Hang Chao. Review and comparison between the wells–riley and dose-response approaches to risk assessment of infectious respiratory diseases. *Indoor Air*, 20(1):2–16, 2010.

[318] Muhammad F. Tahir, Muhammad A. Abbas, Tamkeen Ghafoor, Saima Dil, Muhammad A. Shahid, Mir M.H. Bullo, Qurat ul Ain, Muazam A. Ranjha, Mumtaz A. Khan, and Muhammad T. Naseem. Seroprevalence and risk factors of avian influenza h9 virus among poultry professionals in Rawalpindi, Pakistan. *Journal of Infection and Public Health*, 12(4):482–485, 2019.

[319] Nassim Nicholas Taleb. *Antifragile: Things that Gain from Disorder*. Random House Incorporated, 2012.

[320] Nassim Nicholas Taleb. How much data do you need? An operational, pre-asymptotic metric for fat-tailedness. *International Journal of Forecasting*, 35(2):677 – 686, 2019.

[321] Nassim Nicholas Taleb. Hypothesis testing in the presence of false positives: The flaws in the Danish mask study, 2020. Accessed 2020-12-07.

[322] Nassim Nicholas Taleb. The masks masquerade. `https://medium.com/incerto/the-masks-masquerade-7de897b517b7`, 2020. Accessed 2021-02-04.

[323] Gouhei Tanaka, Chiyori Urabe, and Kazuyuki Aihara. Random and targeted interventions for epidemic control in metapopulation models. *Scientific reports*, 4(1):1–8, 2014.

[324] Biao Tang, Xia Wang, Qian Li, Nicola Luigi Bragazzi, Sanyi Tang, Yanni Xiao, and Jianhong Wu. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *Journal of Clinical Medicine*, 9(2):462, 2020.

[325] Geert R Teisman and Erik-Hans Klijn. Complexity theory and public management: An introduction. *Public Management Review*, 10(3):287–297, 2008.

[326] Peter FM Teunis, Nicole Brienen, and Mirjam EE Kretzschmar. High infectivity and pathogenicity of influenza a virus via aerosol and droplet transmission. *Epidemics*, 2(4):215–222, 2010.

[327] Stefan Thurner, Rudolf Hanel, and Peter Klimek. *Introduction to the Theory of Complex Systems*. Oxford University Press, 2018.

[328] Tommaso Toffoli and Norman Margolus. *Cellular Automata Machines: A New Environment for Modeling*. MIT Press, 1987.

[329] Juliana Tolles and ThaiBinh Luong. Modeling epidemics with compartmental models. *JAMA*, 323(24):2515, 2020.

[330] Brett Trueman. Analyst forecasts and herding behavior. *The Review of Financial Studies*, 7(1):97–124, 1994.

[331] Alan M Turing. Computing machinery and intelligence. In Robert Epstein, Gary Roberts, and Grace Beber, editors, *Parsing the Turing Test*, pages 23–65. Springer, 2009.

[332] Mitsuo Uchida, Minoru Kaneko, Yoshihiko Hidaka, Hiroshi Yamamoto, Takayuki Honda, Shouhei Takeuchi, Masaya Saito, and Shigeyuki Kawa. Effectiveness of vaccination and wearing masks on seasonal influenza in matsumoto city, japan, in the 2014/2015 season: An observational study among all elementary schoolchildren. *Preventive Medicine Reports*, 5:86–91, 2017.

[333] Robert E Ulanowicz. The balance between adaptability and adaptation. *BioSystems*, 64(1-3):13–22, 2002.

[334] Robert E Ulanowicz. The dual nature of ecosystem dynamics. *Ecological Modelling*, 220(16):1886–1892, 2009.

[335] UNESCO. *Encyclopedia of Life Support Systems*. `https://www.eolss.net/`.

[336] U.S. Department of Homeland Security Transportation Security Administration. TSA checkpoint travel numbers for 2020 and 2019. https://www.tsa.gov/coronavirus/passenger-throughput, 2020. Accessed 2020-6-28.

[337] U.S. Department of Transportation. 2018 traffic data for U.S. airlines and foreign airlines U.S. flights. https://www.bts.dot.gov/newsroom/2018-traffic-data-us-airlines-and-foreign-airlines-us-flights. Accessed 2020-3-12.

[338] Pauline Van den Driessche and James Watmough. Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180(1-2):29–48, 2002.

[339] Marianne van der Sande, Peter Teunis, and Rob Sabel. Professional and home-made face masks reduce exposure to respiratory infections among the general population. *PLoS ONE*, 3(7):e2618, 2008.

[340] Alessandro Vespignani. The fragility of interdependency. *Nature*, 464(7291):984–985, 2010.

[341] Victor Virlogeux, Vicky J. Fang, Joseph T. Wu, Lai-Ming Ho, J. S. Malik Peiris, Gabriel M. Leung, and Benjamin J. Cowling. Incubation period duration and severity of clinical disease following severe acute respiratory syndrome coronavirus infection. *Epidemiology*, 26(5):666–669, 2015.

[342] Patrick G. T. Walker, Charles Whittaker, Oliver J. Watson, Marc Baguelin, Peter Winskill, Arran Hamlet, Bimandra A. Djafaara, Zulma Cucunubá, Daniela Olivera Mesa, Will Green, Hayley Thompson, Shevanthi Nayagam, Kylie E. C. Ainslie, Sangeeta Bhatia, Samir Bhatt, Adhiratha Boonyasiri, Olivia Boyd, Nicholas F. Brazeau, Lorenzo Cattarino, Gina Cuomo-Dannenburg, Amy Dighe, Christl A. Donnelly, Ilaria Dorigatti, Sabine L. van Elsland, Rich FitzJohn, Han Fu, Katy A. M. Gaythorpe, Lily Geidelberg, Nicholas Grassly, David Haw, Sarah Hayes, Wes Hinsley, Natsuko Imai, David Jorgensen, Edward Knock, Daniel Laydon, Swapnil Mishra, Gemma Nedjati-Gilani, Lucy C. Okell, H. Juliette Unwin, Robert Verity, Michaela Vollmer, Caroline E. Walters, Haowei Wang, Yuanrong Wang, Xiaoyue Xi, David G. Lalloo, Neil M. Ferguson, and Azra C. Ghani. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science*, 369(6502):413–422, 2020.

[343] J Wallinga and M Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, 2006.

[344] Jacco Wallinga and Marc Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, 2007.

[345] Duncan J. Watts, Roby Muhamad, Daniel C. Medina, and Peter S. Dodds. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences*, 102(32):11157–11162, 2005.

[346] Karl E Weick. *Sensemaking in Organizations*. Sage, 1995.

[347] Carmen Weigelt and MB Sarkar. Performance implications of outsourcing for technological innovations: Managing the efficiency and adaptability trade-off. *Strategic Management Journal*, 33(2):189–216, 2012.

[348] Eric W. Weisstein. Kermack-Mckendrick model. https://mathworld.wolfram.com/Kermack-McKendrickModel.html. Accessed: 2021-04-17.

[349] Chad R Wells, Pratha Sah, Seyed M Moghadas, Abhishek Pandey, Affan Shoukat, Yaning Wang, Zheng Wang, Lauren A Meyers, Burton H Singer, and Alison P Galvani. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proceedings of the National Academy of Sciences*, 117(13):7504–7509, 2020.

[350] Justin Werfel and Yaneer Bar-Yam. The evolution of reproductive restraint through social communication. *PNAS*, 101(30):11019–11024, 2004.

[351] Karoline Wiesner, Alvin Birdi, Tina Eliassi-Rad, Henry Farrell, David Garcia, Steve Lewandowsky, Patricia Palacios, Don Ross, Didier Sornette, and Karim Thébault. Stability of democracies: A complex systems perspective. *European Journal of Physics*, 40(1):014002, 2018.

[352] David Sloan Wilson. *Does Altruism Exist?: Culture, Genes, and the Welfare of Others*. Yale University Press, 2015.

[353] World Health Organization. Coronavirus disease 2019 (COVID-19): Situation Report 72. https://apps.who.int/iris/bitstream/handle/10665/331685/nCoVsitrep01Apr2020-eng.pdf, 2020.

[354] World Health Organization. Novel coronavirus – China. www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/, 2020.

[355] Worldometer. COVID-19 coronavirus outbreak. https://www.worldometers.info/coronavirus/. Accessed 2020-8-28.

[356] Jiang Wu, Fujie Xu, Weigong Zhou, Daniel R. Feikin, Chang-Ying Lin, Xiong He, Zonghan Zhu, Wannian Liang, Daniel P. Chin, and Anne Schuchat. Risk factors for SARS among persons without known contact with SARS patients, Beijing, China. *Emerging Infectious Diseases*, 10(2):210–216, 2004.

[357] Joseph T Wu, Kathy Leung, and Gabriel M Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, 395(10225):689–697, 2020.

[358] Shuangsheng Wu, Chunna Ma, Zuyao Yang, Peng Yang, Yanhui Chu, Haiyan Zhang, Hongjun Li, Weiyu Hua, Yaqing Tang, Chao Li, and Quanyi Wang. Hygiene behaviors associated with influenza-like illness among adults in Beijing, China: A large, population-based survey. *PLOS ONE*, 11(2):e0148448, 2016.

[359] Zunyou Wu and Jennifer M. McGoogan. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of

72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*, 323(13):1239–1242, 2020.

[360] Jingyi Xiao, Eunice Y. C. Shiu, Huizhi Gao, Jessica Y. Wong, Min W. Fong, Sukhyun Ryu, and Benjamin J. Cowling. Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—personal protective and environmental measures. *Emerging Infectious Diseases*, 26(5):967–975, 2020.

[361] Saber Yezli and Jonathan A Otter. Minimum infective dose of the major human respiratory and enteric viruses transmitted through food and the environment. *Food and Environmental Virology*, 3(1):1–30, 2011.

[362] John Zaller. *The Nature and Origins of Mass Opinion*. Cambridge University Press, 1992.

[363] Lijie Zhang, Zhibin Peng, Jianming Ou, Guang Zeng, Robert E. Fontaine, Mingbin Liu, Fuqiang Cui, Rongtao Hong, Hang Zhou, Yang Huai, Shuk-Kwan Chuang, Yiu-Hong Leung, Yunxia Feng, Yuan Luo, Tao Shen, Bao-Ping Zhu, Marc-Alain Widdowson, and Hongjie Yu. Protection by face masks against influenza A(H1N1)pdm09 virus on trans-Pacific passenger aircraft, 2009. *Emerging Infectious Diseases*, 19(9), 2013.

[364] Huijuan Zhou, Chengbin Xue, Guannan Gao, Lauren Lawless, Linglin Xie, and Ke K. Zhang. Characterizing the transmission and identifying the control strategy for COVID-19 through epidemiological modeling. *medRxiv 2020.02.24.20026773*, 2020.