Evolutionary Theory and the Dynamics of Institutional Change *

Henry Farrell (George Washington University)
Cosma Shalizi (Carnegie-Mellon/The Santa Fe Institute)

1 Introduction

How do we explain the dynamic processes that underlie institutional change? This is a crucial — and still unresolved — question for the social sciences. Sociologists have tended to analyze institutional change in terms of agency, focusing on the key role of entrepreneurial actors in shaping institutional outcomes (Fligstein and McAdam, 2011) . Rational choice scholars in political science and political economy have more and more sought to use game theory to explain how institutional change occurs, focusing in particular on institutions as equilibria that arise in infinitely iterated games (Greif, 1994; Greif et al., 1994; Milgrom et al., 1990). Historical institutionalists have sought increasingly to capture the specific mechanisms that guide institutional change over time (Hacker, 2004; Mahoney and Thelen, 2009; Thelen, 2004).

Each of these approaches faces its own particular difficulties. Sociologists' emphasis on institutional entrepreneurialism helps them to capture the genuine creativity that actors can exercise in creating, recreating or extending rules. However, just because of this focus, sociologists have found it hard to identify possible constraints on entrepreneurialism that would really help build a proper theory. Rational choice scholars' concern with equilibria makes it very difficult to identify processes of endogenous change (a game theoretic equilibrium is, by definition, self-reinforcing) and furthermore makes it difficult for them to speak convincingly to the kinds of new institutions that might come into being when an old equilibrium breaks down. Historical institutionalists continue to face difficulties in identifying the relationship between major conjunctural change, and processes of gradual interstitial change.

It is, as always, far easier to identify imperfections in existing work than to propose useful ways in which they might be fixed. Even so, we seek in this article, — perhaps with more ambition than good sense — to provide a new account of institutional change that at least helps to rectify some of these flaws. To do this, we draw upon an extensive literature in *evolutionary theory*. An evolutionary account at a minimum requires both a mechanism of transmittable variation, which generates units with transmittable differences, and a mechanism of selection, which determines the

*We are grateful to Mark Blyth, Ellen Immergut, Elinor Ostrom, Sven Steinmo, and Kathleen Thelen for their comments on an earlier version of this paper. Henry Farrell gratefully acknowledges the support of the Woodrow Wilson International Center for Scholars.

relative success with which these differences propagate. In biological evolution, mechanisms such as sexual reproduction, mutation, and the swapping of genes among bacteria may lead to variations, which are then selected over by natural selection (relevant differences in characteristics will lead to different rates of reproductive success). In the social sciences, as we discuss at length below, the mechanisms of variation and selection may be quite different.

Thus, we propose an *evolutionary* account of the dynamics of institutional change - that is, an account which bases its arguments on clearly defined mechanisms of transmissible variation and of selection among those variations. This account serves both as a general example of the benefits of evolutionary accounts in the social sciences, and as a specific approach to integrating arguments about evolutionary change in institutions with a power based account. In order to explain variation in actions that may lead to institutional change, we draw upon results from cognitive science (while briefly discussing related results from machine learning theory). In order to explain how some actions are selected (and hence may lead to institutional change) at the expense of others, we draw upon both network theory and theories of bargaining power.

We do not, of course, claim that our account offers a complete theory of institutional change. Like all such theories it focuses on some explanatory factors and associated mechanisms at the expense of others. Furthermore, even as we seek to address the problems of existing theories, we import new problems of our own (we discuss some of these in the concluding section). Finally, a proper evolutionary account will have relatively limited predictive use. Much of the action of such an account lies in unobservables. Most particularly, the general contours of the 'fitness landscape' - the possible peaks and troughs in the set of solutions that populations explore across will be unknown.

Even so, our arguments provide an important advance on the existing literature. The value of new explanatory approaches is not only in solving problems that perplex the existing literature, but in identifying new questions and causal relationships that are difficult or impossible to discern using existing theories.[1]

We believe that by applying evolutionary theory, we can indeed highlight different mechanisms and causal relationships that have explanatory power. In Ernest Gellner's description:

> human ideas and social forms are neither static nor given. In our age, this has become very obvious to most of us; and it has been obvious for quite some time. But any attempt at understanding of our collective or individual predicaments must needs be spelt out against the backcloth of a vision of human history. If our choices are neither self-evident nor for keeps, we need to know the range of alternatives from which they were drawn, the choices that others have made or which were imposted on them. We need to know the principles or factors which generate that range of options. The

[1] Thus, for example, the concept of 'transaction costs' which has more or less completely resisted efforts at quantification or measurement despite the best efforts of e.g. **?**, has nonetheless had a transformative impact on our understanding of politics and the economy, by pointing to a set of empirical relationships and associated research questions that would otherwise be invisible.

> identification of those principles or factors is not beyond our capacities, even if specific prediction continues to elude us (Gellner, 1988, p.12).

Like Gellner, we believe that the identification of underlying mechanisms and hence the clearer specification of the range of alternatives available to actors, has clear worth, even without any very great capacity for prediction.[2] It provides us with a better understanding of the options available to actors and collectives in particular situations, even if it does not necessarily tell us which option those actors will fix upon. The next two sections of the article discusses evolutionary theory in greater detail, discussing mechanisms of *variation* and *selection*. The succeeding section lays out the microfoundations of our argument, drawing on both sociology and simple game theory. After carrying out the groundwork, we then advance an evolutionary theory of institutional change that uses cognitive science and network theory. We then assay our arguments by showing how they can both help to resolve key debates over cognitive democracy and the political economy of capitalism as well as suggesting new research agendas.

Evolutionary Theory and the Social Sciences

Evolutionary theory has a decidedly poor reputation in the social sciences. Many social scientists are reluctant to acknowledge that evolutionary accounts *as such* can explain anything important about human social behavior. The publicity given to a noxious sub-literature in evolutionary psychology, whose dedication to crafting crude functionalist accounts of complex human phenomena (Miller and Kanazawa, 2007) is matched only by its incompetence at data analysis (Gelman, 2007), has done much to stroke this hostility. Moreover, many social scientists are rightly skeptical of self-proclaimed evolutionary theories of social phenomena which aim at "revealing a universal and irreversible direction of historical development" (Toulmin, 1972, p. 322), generally by laying down a sequence of progressively more complex (differentiated, advanced, etc.) stages, and supposing that Something moves societies, cultures, institutions, etc. from stage to stage, or stalls them. Such theories are scientifically useless (Boudon, 1986), typically exercises in self-congratulation (Gellner, 1988), and often actively pernicious (Popper, 1960; but see also Bannister 1989; Adcock 2009). Fortunately they have nothing to do with theories of variation and selection in adaptive populations, except for borrowing the latter's prestige and mis-applying some of its rhetoric (Toulmin, 1972, ch. 5). Finally, and for very good cause, overtly functionalist reasoning of the kind often employed by evolutionary theorists has become unpopular in the social sciences (Elster, 1985; Barry, 1988) (although it still lurks, barely concealed, beneath the elegant surface of many technically fluent game theoretic arguments).

Yet these objections surely do not suffice to dispose of evolutionary thinking. First, an evolutionary theory of social phenomena certainly does not have to reduce the latter, via crude adap-

[2]Indeed, if our reader imagines that our intellectual project is to imagine what Gellner's philosophical history might have looked like had he been able to read extensively across contemporary evolutionary theory and related fields, she will not be far wrong.

tationism, to straightforward expressions of simple biological imperatives. Boyd and Richerson (1985, 2005); Richerson and Boyd (2006) for example discuss how the human capacity for creating, expressing and learning from culture is surely a result of genetic selection - but stress how the *specific features* of particular human cultures may change as a result of forces quite distinct from genetic selection. More generally, an evolutionary theory does not need to invoke biological adaptations and fitness in its explanations of particular phenomena (as in Wilson (1999)), nor does it require us to find social analogs for the specific mechanisms of variation, selection and retention which appear in biological evolution (as in (Lynch, 1996)). Evolutionary theory does not imply any kind of substantial progress in biology and surely (despite its occasional rhetorical invocation to the contrary purpose) says nothing about progress in society either.[3]. Finally, there is no reason why evolutionary explanations must necessarily be functional ones. Certainly — as Gould (2002, p. 31) emphasizes — the Darwinist account of genetic reproduction is a functionalist one in which the "organism supplies raw material in the form of 'random' variation, but does not 'push back' to direct the flow of its own alteration from inside." But as we have already noted, there is no reason why social scientists need to confine themselves to examining those mechanisms that have already been investigated by biologists.

So what *does* a proper evolutionary theory of change imply? Very simply, it requires that the theory of change invoke and describe the relationship between mechanisms of transmittable variation and selection.[4] The explanation should first explain how variation occurs, so that objects with observable and interesting differences are produced through some process. It must then explain how selection occurs - that is, how some principle operates to select certain variations and not others, and of how variations that have been selected are reproduced preferentially. Thus, evolutionary biology invokes some mechanisms that explain variation — most obviously mutation and (in sexually-reproducing species) the mixing of genes between gametes. The mechanism of

[3]Some scholars in the social sciences (Hodgson and Knudsen, 2008) have recently advocated very slightly more sophisticated — but ultimately ill-conceived — analogous claims from evolutionary theory about the increase of social complexity over time. If complexity is measured by the internal differentiation of organisms, there seems to have been a long term trend for the maximum complexity in the biosphere to rise over time (Bonner, 1988; McShea, 1996), though of course the overwhelming majority of organisms remain single-celled bacteria. (Measurement is hard to do precisely here, because soft tissues are poorly reflected in the fossil record, and it is far from clear that we could, e.g., infer the existence of the mammalian immune system from a fossilized lab guinea pig.) Theorists are divided on whether this is an indirect consequence of selection for increased size, with allometric constraints necessitating a refined division of labor within larger organisms (Bonner, 1988), or whether this should be expected even under random drift in the space of organisms (McShea and Brandon, 2010). If one is interested in more theoretically-grounded measures of complexity (reviewed in e.g. Badii and Politi 1997; Shalizi and Crutchfield 2001; Wiesner and Ladyman 2010), then even less is known (Crutchfield and Schuster, 2003; Gershenson and Lenaerts, 2008). Some current research (Hodgson and Knudsen, 2008) uncritically adopts the concepts of Adami and collaborators, beginning with Adami and Cerf (2000), which take "complexity" to be (a constant minus) the entropy of the genome — an approach well-understood in the literature on measuring complexity to be quite unsatisfactory (Feldman and Crutchfield, 1998). In principle, complexity measures like those of Shalizi et al. (2004) could be applied to computational models, to biological systems (like the genetic variants of *Dictyostelium* morphogenesis studied in Sawai et al. 2005), or even social assemblages. The practical obstacles would be formidable, however, and it remains more of an aspiration for the future than a resource on which social theory can draw.

[4]Social scientists interested in 'generalized Darwinism' often use a slightly expanded formulation — that it should involve mechanisms of variation, selection and retention. The idea of retention is captured here through the requirement that variation be transmittable.

selection in evolutionary biology is natural selection, where organisms (or individual genes, or perhaps even groups of organisms.[5]) that are better adapted to their environment are more likely to survive, reproduce, and pass their traits along to their descendants. (Sexual selection is natural selection carried out by the environment of choosy conspecifics.)

Many authors, going back to a quite early date, have noted that this pattern of explanation can be abstracted from biology to any domain where there are mechanisms of transmittable variation and of selection. A partial and incomplete list would include: James (1880); Lotka (1924); Toulmin (1972); Dawkins (1976, 1982); Hull (1988); Nelson and Winter (1982); Runciman (1986, 1997, 1998). Within the social sciences, Spruyt (1996) points to mechanisms which produced variation in state form in early modern Europe, arguing that three distinct forms — the sovereign state, the city-state, and the city-league — emerged at roughly the same time. He argues that mechanisms of selection (primarily involving economic efficiency) led to the sovereign state predominating over its rivals, and reproducing itself through conquest and emulation. However, there has been remarkably little attention paid in the social sciences to the burgeoning mathematical literature that seeks to model evolutionary processes. We do not seek to provide a complete overview of this literature here for reasons of space. Instead, we identify commonalities in how these differing modeling techniques capture the dynamics of evolutionary change.

What is most important is that these models view (or at a pinch in some cases, can be treated as viewing) evolutionary change as a process of search. More specifically, evolution involves search carried out across a given population for relatively efficient solutions to problems posed both by the underlying environment and by the population itself. One should be quite clear that the term 'efficient solution' carries no particular normative weight - efficiency at any point of time will be local (there may, and very likely will, be more efficient solutions elsewhere in the space being modeled), and its specific form will be determined by the mechanism of selection (which may involve principles that are retrograde in terms of morality, social wellbeing, economic efficiency or any other reasonable normative measure that one might care to apply).

This allows one to capture the importance of mechanisms of selection and variation in a generally useful way. The mechanism of variation throws up possible solutions, while the mechanism of selection determines which of these possible solutions will propagate, and which will not. It also allows one to reach broad conclusions about the circumstances under which change will or will not take place.

Two complementary approaches to thinking about this are particularly helpful: one is to regard evolutionary processes in large populations as a particular kind of stochastic dynamical system, and the other is to consider them as processors of information.

The first, stochastic-dynamical approach examines how the rate of variation and the relative

[5]The question of group selection is notoriously controversial within biology; fortunately, we do not need to enter into it here. We will just note that while naive forms of group selection cannot, indeed, survive the criticisms of Williams (1966) and related authors, there are at least models, with fully explicit microstructure, within which group-level effects occur (Pepper and Smuts, 2000, 2002) through mechanisms other than kin selection, though there is some dispute as to whether they should be said to exhibit "group selection". See Bowles (2004); Bowles and Gintis (1998, 2004) for applications of such models to human groups, and Gould (2002).

difficulty of finding paths from lower fitness to higher fitness 'states' across a given population intersect to affect the likelihood of change. Here, recent work in theoretical biology and evolutionary games has lead to qualitatively similar results for a broad range of evolutionary models (Dupuis, 1988; Benaïm and Weibull, 2003; Nilsson and Snoad, 2000, 2002b,a; van Nimwegen et al., 1997; Young, 1998). The common element of these results is that evolutionary processes obey *large deviations principles*, under which the probability of a certain historical trajectory $x(t)$ of strategies in the population is roughly exponential, $\approx \exp\left\{-\epsilon^{-1} \int_{t=0}^{T} L(x(t))dt\right\}$.

The two most significant terms within this approximation are the scaling factor ϵ, which reflects the noise level of the evolutionary process, and the integral within the exponent — which represents the "action," or relative difficulty of the trajectory (low action trajectories are relatively easy to find, while high action trajectories are relatively hard). ϵ reflects the consequences of the mechanism of variation for change across a given population. It grows with the frequency and magnitude of variations, and shrinks with the population size (the larger the population, the more difficult it will be for a given variation to spread across it).[6] The action function reflects the consequences of the mechanism of selection — but also the size of the selectively-neutral set of alternatives that must be explored before a transition to a higher-fitness state (van Nimwegen and Crutchfield, 2000).

The action function not only reflects the mechanism of selection, but also the composition of the population of possible variations over which this mechanism selects. If there are many selectively-neutral variations, and few selectively-superior variations, then it will take longer on average to move to a higher fitness state than it would if there were fewer selectively-neutral variations and more selectively superior ones.

More generally, the large deviations principle implies (Freidlin and Wentzell, 1998) that the typical time it takes a population to move from one state x to another y is proportional to $\exp\{-V(x,y)/\epsilon\}$, where the "potential" $V(x,y)$ reflects the cost of the least-action path from one state to another. In other words, when the rate of variation is low, the time needed to move the population from one state to another can be exponentially long. Substantively, this reflects the fact that it takes a long time for the necessary deviations to appear in the population, and, having appeared, to become established. The action function itself reflects not just selective effects, but also the size of the selectively-neutral set of alternatives which must be explored (the "neutral network"); once the right variants have appeared, the actual transition can be comparatively rapid. More simply expressed, some possible variations will have no significant consequence for evolutionary fitness. They will be either no better than, or inferior to the existing state. If there are many such possible variants within the space to be explored, and few selectively superior variants, then on average it will take a long time for the latter to be found.

The other useful set of ideas from evolutionary dynamics concerns the role of information in

[6]Note that this means that the most probable trajectories of the population are the ones which minimize the action $\int L(t)dt$, and with small ϵ these trajectories become overwhelmingly more likely than others. (There may be multiple action-minimizing trajectories, in which case the large deviations principle does not select between them.) In plainer language, the principle suggests that when there is a low level of background variation, only 'easy' paths are likely to be found. This least-action principle does *not* require that any individual or class of individuals in the population try to minimize the action, or even that their own behavior is governed by an optimization principle.

the process of selection and population change. A process of evolutionary search can be regarded performing a kind of distributed information processing, in which the signal of selective events provides information about the search landscape.[7] "Information" here is to be understood in the sense which Shannon introduced into communications theory (Shannon, 1948; Cover and Thomas, 1991), but which is perfectly applicable despite the fact that evolving populations are not, of course, artificial coding systems Bergstrom and Rosvall (2009). Information about the fitness function comes to be stored in the population distribution; the rate at which such information can be introduced into the population is limited by the Shannon entropy (roughly, *ex ante* uncertainty) in which members of the population will be selected (Bergstrom and Lachmann, 2005; Kussell and Leibler, 2005; Rivoire and Leibler, 2010). Moreover, the rate at which the average fitness increases is limited by the existing dispersion of fitnesses. When this is low, because fitness differences are small, selection has very little ability to re-shape the population and move it to higher fitness parts of the landscape. A version of such a result was proved long ago by Fisher for a specific class of models of biological evolution (his "Fundamental Theorem of Natural Selection"; Fisher, 1958), and is implicit in the definition by Haldane (1954) of the quantity of selection, but the recent work extends this considerably. This is connected to the results mentioned earlier on search dynamics, as situations of low information input from selection correspond to exploring selectively-neutral regions, while transitions from one selective-neutral region to another create exploitable fitness differences.

We note that most of these findings stem from evolutionary models in biology rather than the social sciences. However, given their consistency with a wide variety of models within biology (they do not specifically invoke e.g. sexual reproduction or other specifically biological micromechanisms, and can be derived from models of 'quasispecies'(Nilsson and Snoad, 2002b,a) which bear a pronounced resemblance to some operational definitions of institutions in the social sciences (Ostrom and Basurto, 2009)), we do not believe that it is at all a stretch to argue that they will apply *pari passu* to evolutionary accounts within the social sciences too.[8] Demonstrating the usefulness of these broad findings still requires specific arguments about the mechanisms of evolution in specific areas of the social sciences. We turn to this task in the next three sections.

Both of these approaches capture aspects of change that are important for institutional theory. A large body of work seeks to determine the timing and circumstances of institutional change, while a burgeoning literature examines the relationship between institutional change and informational flows. As we discuss below, these approaches provide us with new perspectives on both of these questions. Before doing this, however, it is necessary to provide a more specific evolutionary account of institutional change. We turn to this task in the next three sections.

[7] In fact, the supposedly-normative process of Bayesian learning by updating a posterior distribution is a limiting special case of such evolutionary search (Shalizi, 2009).

[8] However, in the absence of fully-specified mathematical models of the evolutionary process in question, it is difficult to be as precise as we would like about this.

Cognition and Strategy

The previous section sets out some of the broad advantages of evolutionary accounts for the social sciences. We now turn to building the foundations of an evolutionary account.

We start from the premise, best explored by Gellner (1988), that both cognition and coercion are crucial to processes of institutional change. This leads us to combine arguments from strategic logic and from cognitive science. There are disadvantages as well as advantages to our choice of microfoundations. The flaws of game theory are perhaps too widely discussed to require further rehearsal. We seek to minimize these disadvantages by situating our strategic arguments in a broader cognitive framework, by using only very simple game theoretic arguments, and by specifically eschewing the claim that actors necessarily share common knowledge. Hence, we use the language of game theory to emphasize how strategic considerations will affect actors' understanding of each other's likely actions and their own best replies, rather than e.g. as the basis for claims based on sophisticated equilibrium concepts. Under our account, actors' internal representations of the strategic situation that they find themselves in will be embedded in a broader set of cognitive understandings. Moreover, we treat these cognitive factors as analytically prior to the strategic ones. The former lay out the menu of possible actions that actors can conceive of, while the latter tell us which of these menu items actors will choose.

To capture the consequences of cognitive factors, we look to individuals' schemas. Schemas, as we use the term here, are organized representations of the world and its relationships that allow people to categorize situations as belonging to certain types, and tell people what the possible actions are in the relevant type of situation. While a schema may help individuals identify various actions that are possible in a given situation, it will not tell the individual which of these actions is the *right* one to adopt. Schemas focus our attention on specific aspects of reality, foreclosing some possibilities and disclosing others, but they do not dictate which of the disclosed actions one should take.

Ariel Rubinstein (1991) in his discussion of the interpretation of game theory, discusses the importance of 'language' to strategic choice in ways that help illustrate what we mean. In a charming example, he talks about how he encouraged his infant daughter to choose blocks of different colors, but was nonplussed to discover that she did not consistently choose one color. He swiftly realized that rather than selecting on color, she was selecting on position — she always chose the leftmost block regardless of which color it was. Two points emerge from this that are important to our later discussion of the relationship between schemas and strategic choice. The first is that the "language" of block categorization (or, as we might dub it, the schema of block categorization) is analytically prior to the actual choices that one makes. One makes choices along a particular dimension which is suggested by the relevant schema. In more complicated situations than the one that Rubinstein describes, the schema may not only provide a means for focusing on a particular dimension of choice, but actively obscure some choices and highlight others.

The second is that people operating with radically different schema are likely to have initial difficulty understanding each other, but that even so, differences in schema may be heuristically

fruitful. If one is operating within a schema where color dominates block choice, then one is likely to opt for a heuristic (such as "always choose blue"), which is color dependent. If, one is operating within a schema where position dominates, one is likely to opt for a different heuristic, such as "always choose the leftmost block." And if one has a schema that emphasizes color, but wants to coordinate with someone whose schema emphasizes position, one is likely to have a frustrating time of it until one figures out what is going on.[9] However, exposure to different schemas can have beneficial cognitive consequences and reveal new possibilities of action. If one actor analyzes a situation in terms of one schema, and another in terms of a very different schema, they may — assuming that they can communicate with each other or otherwise learn from each other — be able to combine schemas so as to reach solutions that would otherwise be invisible to them (Page, 2007). We develop this claim further in the next section.

While schemas provide actors with a list of possible actions, they do not dictate which specific actions actors will choose from that list. We argue that each actor will choose the action (while game-theoretic strategies may involve a sequence of iterated actions in multi-stage games, we prefer to concentrate on very simple games here) that she believes to be most advantageous in terms of individual self interest, regardless of its implication for others. Hence, our framework emphasizes individual gain rather than other-regarding concerns. This has implications for our broader argument, which we develop below (most obviously, it sets the stage for an account of institutional change that will emphasize asymmetric bargaining rather than this or that variety of functionalism). However, it is important to note that game theory provides us with no more than a useful shorthand for capturing the relevance of self-interest and individual perceptions of the strategic situation that they find themselves in in a reasonably consistent fashion. Not only do our conclusions fail to imply e.g. actors' ability to solve complex games through backwards induction, but we explicitly decline to use common knowledge as a basic starting assumption. Individuals' understanding of the strategic situations that they find themselves in and of other actors' understanding of that situation and of their own understanding etc will vary in empirically important ways. Not only may actors not share common knowledge of a situation - they may not even agree on what the salient features of that situation are. While, as we discuss below, actors may reach something approximating equilibria - reasonably stable and self-reinforcing understandings of their situations, of how others may be expected to behave etc - these equilibria will be contingent and vulnerable to disruption when new understandings emerge. In effect, they will be islands of temporary stability in a broader evolutionary process - fitness states which (depending on the factors discussed in the previous section) may give way to new fitness states when new paths are explored.

We specify the relationship between schemas and strategic action more precisely by imagining the world of social possibilities as a fitness landscape, each point of which is associated with a pair of actions that can be taken respectively by members of two classes, *Alpha* and *Beta*. These different points have payoff sets with different distributional consequences — at some points, actors

[9]Shared languages or schemas underlie the various mechanisms of salience that Schelling (1960) identifies; see further Schiemann (2000).

belonging to one class do well, while actors belonging to a second class (or classes) do badly; at others, actors belonging to the second class do well, and actors belonging to the first class do badly; at others yet, both receive roughly symmetric payoffs and so on.[10]. Nor is this a simple 'divide the pie' game — the overall payoffs vary from point to point. In other words, this landscape is rugged in two senses — both the local optima for individuals belonging to particular classes (the points at which the payoffs accruing to members of these particular classes are higher) and the collective local optima (the points at which the collective payoffs for all actors are higher) vary substantially.

This landscape is similar to the ones described by Scott Page (2007) and other social scientists who are interested in the ways in which diversity of perspectives (or, in our language, schemas) can help agents work together to identify global optima. However, there is a key difference — the points on our landscape have different distributional implications for different classes of actors. Actors belonging to one class might prefer one set of optima; actors belonging to another class would likely prefer a different set, that will probably overlap only somewhat or not at all. Hence, coordination is far more fraught with difficulty than in Page's models of cooperative problem-solving.[11]

Now, we inhabit our landscape with actors. As in Page's work, actors' ability to perceive particular optima will depend on their schemas. However, unlike the actors in Page's cooperative models, social actors in this world are not engaged in the common endeavor of finding a global optimum that, say, maximizes overall payoffs. Actors will prefer points that maximize the payoffs accruing to their particular class over equilibria that maximize overall payoffs but that have lower specific payoffs for their class. (Said another way, each class of actor has its own landscape, imperfectly correlated with the other's.) Schemas will only reveal a quite limited subset of the range of possible actions to actors, and hence a tightly constrained set of points on the fitness landscape. Individual human beings' cognitive capacities, and hence their ability to perceive the world will be constrained by their schemas. Because they can only perceive certain actions, they can only perceive certain peaks, and not others. While actors will want to converge on points that are locally optimal (for themselves), they do not know whether these points are globally as well as locally optimal, or whether, alternatively, there are other potential points on the landscape that would be better if only they could perceive them and take the appropriate actions (and get others to take the appropriate actions too).

Within these cognitive constraints, we may expect actors to behave strategically. They will prefer outcomes that increase their payoffs, and will use their bargaining strength to try to achieve these outcomes. In many social situations, different classes of actors will systematically differ in their bargaining strength — that is, actors from one class may be expected to have systematically better bargaining positions, given their available actions, when they are in this situation than actors from another class. This will give rise over time to expectations and institutional rules that will lead actors to converge on a point where the distributional payoffs reflect the asymmetries in bargaining power. Thus, institutional outcomes will reflect both cognitive constraints (the payoffs

[10]For simplicity, we divide the inhabitants of this world into just two classes of actors, but there is nothing important in our argument that hinges on this.

[11]See however Page (2008).

being bargained over are limited to those disclosed by actors' schemas) *and* strategic bargaining (the prevailing institutions will instantiate beliefs that reflect strategic behavior and differences in bargaining strengths). While iterated bargaining and associated communication may sometimes lead to convergence in actors' schemas over time, full convergence is not required for our arguments to hold. It is entirely possible that actors may understand the points that they have converged on in very different terms, as long as each actor's particular understanding and expectations of the other leads her to behave in a somewhat predictable way.[12]

This conducts towards an account of institutions based on strategic as well as cognitive considerations. We understand institutions as rules (implicit or explicit) based on roughly shared expectations about what different classes of actors can be expected to do in various types of social situations.[13] For example, a simple institutional rule, governing two classes of actors, Alphas and Betas, might dictate that when Alphas and Betas interact in situation X, Alphas will play strategy x_1 and Betas will play x_2. There is no reason to presuppose that such institutions are fair, or (in any normatively interesting sense) efficient. The institution discussed above may be advantageous to Alphas, while not necessarily giving Betas much more than they would get from a breakdown of cooperation, and still be sustainable (Knight, 1992).

One historical example of such an institution was the pervasive rule in many parts of the post-Civil War South that African Americans, if walking on the pavement, had to step off the pavement to give way to whites coming in the opposite direction. This institution clearly advantaged white Americans and disadvantaged African Americans. It did nothing to promote efficiency, but did reflect power relations. African Americans who refused to step off the pavement were very well aware that they could suffer drastic consequences in the case of breakdown, including organized collective violence sanctioned by local authorities. Thus, even if they did not consider the institution to be *legitimate*, they had strong reason to obey it. Those who did not obey, because they sought to evade the rule or because they did not understand it, suffered horribly. This — as with other such institutions — created strong expectations about the ways that whites and African Americans should behave towards each other across a variety of situations, conducting towards outcomes that favored white people and disfavored African Americans. Some might object that this institution was too harsh to be representative. They would be completely wrong: it is representative of a wide array of institutions over history which have instantiated serious power inequalities and relied on the threat of extreme physical violence to encourage compliance. Arguably, institutions which do not instantiate such inequalities have been the exception rather than the rule for most of recorded human history.

These arguments about cognition, strategic considerations and institutions lay the groundwork

[12]Put differently, institutions require that actors share reasonably robust expectations about what they and others will do in particular sets of circumstances. They do not require that these expectations be based in commonly shared theories of motivations. Even if I am mistaken in my reasoning as to *why* you comply with a particular institution, and you are mistaken about my motivations, the institution may still endure, as long as both of us continue to comply.

[13]Our understanding of institutions as involving shared, coordinated expectations comes from game theory (Lewis, 2002), but it has older roots in the social sciences (e.g., Hayek 1937), and even the humanities, as in the insightful and suggestive discussion of institutions, expectations and tradition in Hodgson (1993).

for our broader argument in the next section. Actors (in our account) will behave strategically, in that they will seek, on the basis of their understanding of the options available to them, and the motivations of others, to further their individual self-interest. However, their understanding of those options will depend on their schemas - the cognitive frameworks which provide them with an organized representation of the world and its relationships. Not only are individuals' choices dependent on their schemas, but these schemas need not be fully consonant with each other — indeed we can confidently predict that they will not be fully consonant with each other (Sperber, 1996), so that some rough approximation of consensus is the most that can be expected.

Furthermore, institutions will reflect differences in bargaining power between classes of actors (Knight, 1992; Bowles and Naidu, 2008), rather than, say, the efficiency considerations embraced by much of the new institutional economics (Eggertsson, 1990). If actors were fully aware of all the options available to them, and of the state of the world, then we might expect that institutions would perfectly reflect power relations. Whenever changed external circumstances re-balanced the bargaining strength of different classes, institutions would change too. In real life, of course, actors lack full awareness of the world and their options. Institutional change will thus be a more uncertain process, in which learning plays a role alongside power. We describe the relevant mechanisms in the next two sections.

An Evolutionary Account of Institutional Change I - Mechanisms of Variation

Which mechanisms of variation and selection will lead to institutional change in the world described above? In our account, the important unit of variation is the *action* rather than the institution itself. We claim (1) that actors occasionally deviate from these institutionalized expectations, behaving in unexpected ways; (2) that some of these deviations are adopted by others and so diffuse across the relevant population, i.e., are "selected for"; and (3) that successful deviations lead to changes in actors' expectations, and hence to institutional change.[14]

To resume our example from the previous section, if most members of the broader population expect that when Alphas and Betas find themselves in situation X, the former will do x_1 and the latter x_2, we may describe this set of expectations as an institution. However, actors will not always behave as they are expected to. For example, actors may deviate because they do not fully understand an institution, or alternatively wish to test it to see if it still applies, or perhaps want to try to stretch it in unexpected ways. Thus, there will sometimes be deviations, in which actors do not respond to the situation with the expected action.

Such deviations are the key *source of variation* in our account. Here, the 'deviation rate,' is analogous to the 'mutation rate' in evolutionary biology, and is represented by ϵ in the large deviations approximation described previously. There is excellent reason to believe that the deviation

[14]We certainly do not claim that ours are the only possible or appropriate mechanisms one might use in an evolutionary account. See the conclusions.

rate for a society will never be zero. As Balkin (1998, p. 93) puts it, "culture is not a top-down network, in which a single server transmits identical copies of a software upgrade to the various nodes". (A Durkheimian model of culture *would* act like this, locating the collective consciousness of North American society somewhere near Redmond, WA.) Rather, schemas and other cultural representations must be acquired through ordinary processes of learning, which will produce variation, both because of intrinsic limitations of learning, and because of the biology of human memory.[15]

We can see no very straightforward way in which we can generalize about e.g. cross-societal differences in the deviation rate resulting from learning processes or the vagaries of human memory (this is not to say that there *is* no such way; merely that we cannot readily see how to capture it). However, there is a third source of variations, which will plausibly differ from society to society in ways that are easier to capture. The degree of *heterogeneity* in schemas across a particular population of actors is likely to be an important source of new variations - hence we focus on it to the exclusion of other possible sources of variation. We argue that three associated mechanisms (which are loosely associated with sociological, rational choice and deliberative accounts of institutional change) will lead to an association between increased heterogeneity and increased generation of new varieties of action.[16]

First, we may expect that increased heterogeneity in schemas across a society will mean that actors will lead to more frequent disagreements about how to interpret an institution. Sociologists such as (Dobbin, 1997) emphasize how individuals in different societies may have fundamentally different ways of understanding the world, with consequences for their actions. These arguments may also apply, with greater or lesser force, to individuals with different backgrounds within a given society. The greater the heterogeneity in schemas, the more likely that all actors will fail to see a given situation as a particular kind of situation falling under a specific institution. These differences may lead actors to initiate actions that they expect will lead others to respond in a particular way, only to see their expectations confounded. They will also make actors more likely to respond in

[15]These points deserve amplification. The most favorable case for learning without downloading is imitation learning with a teacher providing explicit feedback to the pupil. Even here, however, fundamental results in statistical learning theory show that it is generally not possible for the pupil to *exactly* duplicate the teacher's internal rule on the basis of any finite number of training examples (Watkin et al., 1993; Vapnik, 2000; Engel and Van den Broeck, 2001; Vidyasagar, 2003). As Vapnik (2000) emphasizes, this is simply an ill-posed inverse problem. At best, pupils may be guaranteed to give "probably approximately correct" responses to future situations with the same statistical properties as the training cases. This point applies not just to schemas for social institutions but other forms of learning; Niyogi (2006), for instance, uses it to explain the historical development of languages. The mathematical obstacles in the way of inducing schemas from examples would produce schematic variation in *any* population of learners. (Without getting too science-fictional, think of organizations trying to learn from experience.) The nature of human memory, however, creates an additional source of variation, as has been emphasized by writers such as Sperber (1996) and Turner (1994). Drawing on a tradition in psychology going back to Bartlett (1932) at least, and if anything emphatically confirmed by more recent developments (Schacter et al., 1998), these authors emphasize that our memory does not so much *reproduce* what we have previously learned or experienced as it *reconstructs* it on the basis of currently available information, concepts, emotional states, etc., as well, of course, as actual memory traces. This is, naturally, a further source of conceptual and schematic variation.

[16]Here, we follow Johnson (2002) in understanding "mechanisms" as structures that are not usually observable, that regularly connect causal forces to their effects, and that provide the cogs for a broader social theory. Cf. the philosophical treatment in Salmon (1984), and the connection drawn by Morgan and Winship (2007) with Pearl (2009)'s "front-door criterion" for unconfounded causal inference.

unexpected ways to actions initiated by others. Such 'mistakes' are a potential source of variation.

Second, we may expect that the greater the heterogeneity in schemas across a society, the greater will be the scope for individuals strategically to misinterpret institutions for their own ends. Here, heterogeneity will generate ambiguity, which will in turn offer strategic opportunities to actors to push ambiguities in ways that favor them, and their own interests, perhaps at the expense of others. As Thelen (2004); Hacker (2004); Hall and Thelen (2009) and others have observed, institutions are regularly contested as they are implemented, by actors who wish to reframe these institutions in congenial ways. This is especially so because institutions involve the application of abstract rules to concrete situations, which may not precisely fit under the institution's definition (Hart, 1994). (In statistical learning theory, the analogous problem is known as "data-set shift" (Quiñonero-Candela et al., 2009).) Heterogeneity in schemas may hence provide strategic actors with a valuable resource, making it easier to respond with unexpected actions without appearing directly to dissent and hence making it easier to test the boundaries of institutions (one might perhaps dub this the 'Good Soldier Svejk' mechanism).

Third, the more heterogeneity among schemas there is, the more likely it is that individuals will be able to generate genuinely new insights by communicating among each other and combining the logics of different schemas to new ends. As Page (2007) and others have suggested, greater cognitive diversity makes it much more likely that individuals who encounter others with different points of view, and work together with them, can identify possibilities of action that might otherwise have eluded them. Page's work suggests that intellectual exchange between people with different schemas can lead to enormous gains in the absence of strategic considerations. Even if actors are entirely self-centered, there will likely be some areas where they can communicate to common purpose and coordinate accordingly. Deliberation among individuals with different ways of viewing the world can lead to the discovery of new actions and new combinations of actions.

These new actions may plausibly affect actors' perceived options in either of two ways - through changing the value of actions that are already available, or by changing the set of available actions. The more straightforward way is through *change in the expected value of outcomes.*[17] Consider what happens when, for some exogenous reason, the expected value of fallback options in the case of breakdown changes — say (without loss of generality) that Betas get more or better options than they had previously, while the value of options available to Alphas either remains constant or shrinks. Such changes may happen through, for instance, shifts in relative prices. For example, many scholars hypothesize that the labor shortages resulting from the Black Death increased the bargaining power of serfs vis-a-vis feudal lords in the medieval era, leading to the abandonment of serfdom across much of Western Europe North and Thomas (1973). Since there were fewer peasants, they were less substitutable for each other. Hence, the breakdown option of harsh punishment for

[17] One should bear in mind, although we do not treat this further, that changes to the value of a strategy include its impact on what strategies are available in *other* situations and their values. These cross-dependencies may arise through environmental side-effects, retaliation, reciprocity and reputation, budget constraints, cognitive constraints (as in Bednar and Page 2007]), etc. Hodgson (1993, ch. 4) and Gellner (1988, ch. 2 and *passim*) suggests that the tendency to regard different spheres of behavior as self-contained domains in itself facilitates innovation, by reducing such cross-dependencies, or at least encouraging actors to ignore them.

peasants was less attractive to lords, making them less indifferent to the risk of breakdown. Hence in turn, peasants could better resist the encroachments of their lord when they discovered that breakdown was more unattractive to lords than it had been before. The process through which this transition occurred likely involved much trial and error, under which e.g. lords found through experience that punishing peasants too harshly hurt their interests, and peasants began to push the boundaries of existing institutions through persistent and quasi-surreptitious experimentation.

The second, which can lead to more profound changes, is *change in the repertoire of actions*. Deviating actions may reveal new possibilities of action to actors who carry them out or who observe them, hence reshaping their schemas. In other words, some deviant actions will be *novel*; they did not previously form part of the set of actions that were thought available in a particular situation, or perhaps at all.[18] Deviations like this can be fruitful when they reveal new possibilities of action that are attractive to at least one set of actors. Again, many such actions will prove to be useless, and will not become part of the repertoire of any actor. Such actions will be drawn from the 'neutral set.' But other such actions may reveal new and hitherto unknown points on the fitness landscape (as other actors respond to the new action, using existing or novel actions of their own).

To sum up, we suggest that institutionalists should look to sources of variation in *actions* in order to understand the micro-mechanisms of institutional change, and to heterogeneity of actors' schemas to understand the sources of differences in the 'deviation rate' across societies. Variations in action will be causally linked to heterogeneity of schemas through three associated mechanisms - disagreement in how to interpret institutions, strategically motivated efforts to change the application of institutions, and generation of new insights into possible action through intellectual exchange between individuals with diverse points of view. This suggests that we may expect that societies with higher heterogeneity among schemas will produce more variations in action than societies with lower rates of heterogeneity. In the next section, we turn to issues of *selection* examining which of these actions is likely to be selected (i.e. disseminated so widely across a community that it becomes the expected action in a given situation, hence leading to institutional change) and which will fail to propagate.

An Evolutionary Account of Institutional Change II - Mechanisms of Selection

As we have noted above, not all variant actions will prosper (we may usually expect that only a very small minority will propagate and spread through a given population). Institutional change will happen when when nearly all actors in the population come to view a coordinated set of actions, which was previously perhaps seen as deviant, as the expected actions in a given situation. To a

[18]Logically, it is also possible that actors might *lose* the capacity to follow a previously available strategy. While technological regress certainly happens (Richerson and Boyd, 2006), we cannot think of examples where it eliminated institutionalized strategies, *except* in cases of exogenous change in resource availability, which can be assimilated to changes in the value of existing options, as above.

reasonable first approximation, this will happen when the deviant actions become generalized across the population in question. What explains which variations spread across the population, leading to institutional change and which do not? While deviating actions will obviously be unlikely to spread if they make the actors carrying them out worse off, this is hardly a very selective criterion. For a better understanding of how actions do or do not spread, we (a) draw on contagion models across networks, and (b) treat power asymmetries between actors as the key factor explaining the resistance to infection of nodes within these networks.

Specifically, we treat the spread of actions across a population as a form of diffusion[19], similar in many ways to viral contagion.[20] This allows us to draw on a body of work which treats social learning as a kind of contagion, and examines how it is affected by different kinds of network structure (Newman, 2002, 2003; Durrett, 2006; Kenah and Robins, 2007; Draief and Massoulié, 2010; Newman, 2010; Easley and Kleinberg, 2010).[21] The central problem addressed by this theory is when a diffusing behavior will take off to spread over a non-trivial fraction of the network. The answer depends on whether each individual who adopts the behavior causes, on average,more than one other member of the population to adopt it. This in turn depends both on the probability of transmitting the behavior across any given link, and on the number of "susceptibles" each non-adopter is linked to. When the expected number of new adoptions is strictly greater than one, the number of adoptees will tend to grow exponentially (until saturation sets in).[22] These models tend to have relatively simple accounts of learning, as the straightforward and faithful reproduction of a behavior that has been acquired from others, scamping on the complexities discussed by Sperber (1996) but provide a nuanced account of how learned behaviors (in our case deviant actions) may spread through a population. [23].

Here, the crucial mechanisms of selection will determine whether or not an action is diffused

[19]Note that there is an alternative - to model this spread as a kind of social cascade, in which the probability that a particular node will take up an action will depend on its estimation of whether enough other nodes in its neighborhood are likely to do the same. See Chapter 19 of Easley and Kleinberg (2010) for a useful overview. Such accounts have been used e.g. to model collective action. Given our assumption that only already-successful actions are diffused, a contagion model seems more appropriate for our specific purposes.

[20]The analogy between undesired innovations and contagious diseases seems to have occurred repeatedly to hostile observers. The oldest example we have found is Pliny the Younger calling Christianity a "contagious superstition" (*superstitionis contagio*) in a letter to the Emperor Trajan in 110 (*Epistles* X 96.9).

[21]Since (as discussed below) ties tend to form between nodes with similar attributes, it is always possible that apparent diffusion across a social network is really due to similar individuals reacting similarly to a common external cause (Sperber, 1996). Disentangling this from contagion or social influence raises tricky statistical issues, and often requires heroic assumptions (Shalizi and Thomas, 2010). But no one disputes that social influence exists.

[22]More technically, the theory approximates the early stage of the diffusion of the behavior to a branching process, and the crucial question is whether the branching process is sub- or super- critical. See Newman (2010, ch. 17) or Easley and Kleinberg (2010, chs. 19 and 21) for introductions, and Durrett (2006); Draief and Massoulié (2010) for detailed treatments.

[23]This is a crucial stage of institutional change, though it rarely gets sustained attention in political science and political economy. If institutions involve commonly understood rules and expected behaviors across a given population of actors, then we really need to understand how rules and strategies are disseminated through the population. p.214 Knight and North (1997) note the need for an institutional theory that not only tells us how "how human beings learn and meld beliefs and preferences to reach decisions and hence [explains] the choices that underlie economic theory, but also how and why they develop theories in the face of pure uncertainty, what makes those theories spread among a population and die out, and why ... human beings believe in them and act on them." This article can be understood as an effort to build just such a theory

across the relevant social network, hence becoming generalized as an institution. The key relationship here is the interaction between (a) the topology (the shape of the distribution of links across the various nodes) of the network, and (b) the extent to which actors are more or less likely to resist a new action. The first determines whether or not actors are likely to be exposed to a new possible action. The second determines whether they will take it up, and perhaps spread it further.

Three aspects of network topology are particularly important to diffusion: density, transitivity, and degree variance.

The implications of network density for diffusion are fairly obvious. Increasing the average number of connections between nodes, all else being equal, increases the number of susceptible nodes to which the behavior can spread from those who have already adopted it. This makes diffusion faster and less likely to die out.[24] The sparser the network, the more attractive the new behavior must be in order to achieve a given likelihood of spreading. Increasing the average number of links between nodes will tend to increase the speed and success of diffusion of new best replies. If a network has a low number of average links, we expect that diffusion will be slow and uncertain. Some nodes in the network may be nearly isolated from other nodes, or form clusters which have only tenuous links to the broader network. Innovations will tend not to disseminate to, or from, these nodes. However successful in its own terms, an innovation may simply peter out — or (in societies where powerful actors can draw upon resources of organized violence) be stamped out before it diffuses widely. In contrast, it will be harder to keep innovations from spreading rapidly across dense networks. It is very likely that all parts of the network will be well connected to each other so that there will be few or no "structural holes" (Burt, 1992). Even with organized violence, it will be hard to damp down or prevent the spread of new behaviors, as in no society are professional providers of violence capable of disciplining all, or even most individuals, should those individuals simultaneously begin behaving in undesirable ways (Hume, 1994, ch. 3, "Of the first principles of government").

Transitivity may also have important consequences. A network is *transitive* if, when nodes A and B are both connected to C, then nodes A and B are disproportionately likely to be directly connected. (This is also called *triadic closure* or *clustering*.) Transitivity is ubiquitous in social networks, not least because it follows from the even more common phenomenon of "homophily" or "assortative mixing", the tendency for links to form between nodes because they share similar attributes (McPherson et al., 2001).[25] Networks with high transitivity tend to consist of tightly connected clusters of nodes, also called "communities" or "modules", which are much more weakly tied to other clusters. This affects diffusion in two ways. High transitivity means that once a behavior appears within a cluster, it is more likely to spread and become established within that cluster. At the same time, however, the more clustered a network is, the *less* likely it is that a behavior will diffuse across networks to take over the network as a whole, because links between

[24] Duncan Watts (2004) suggests that innovation becomes less likely again in very densely connected networks — but the learning model underlying his account of diffusion is less likely to apply to the kinds of acquired behaviors we discuss here.

[25] If A and B are both similar to C, then they are similar to each other.

clusters are sparse.[26] The first node in a cluster to adopt the innovation, say A, finds many susceptible neighbors, and so initially the innovative can spread rapidly. However, when a node B adopts the innovation from A, transitivity implies that B's neighbors are also likely to be A's neighbors, and so no longer susceptible. After a fairly short amount of time, there are few susceptible nodes left within the cluster, and few links outside the cluster, so the innovation becomes, as it were, endemic but not pandemic. High transitivity networks will tend to see quick diffusion of new behaviors within particular clusters, while spread between clusters will be slow, sporadic, or even absent. Lowering transitivity will make it more likely to see more rapid diffusion of advantageous — and hitherto unknown — actions across the entire population.[27]

Finally, the higher the variance in the number of links associated with particular nodes (their "degree"), the easier it is for innovations to diffuse across the population. High variance implies that some nodes will have especially large numbers of links to other nodes, which may include non-local nodes. In networks with high variance in the number of links between nodes (e.g. networks with scale-free or lognormal distributions of links) the most linked nodes play a very important role in determining whether behaviors are disseminated or not. If these nodes pass on a behavior, then it is far more likely to be disseminated widely across the network. If these nodes do not, they may play a very important role as firebreaks. In other words, if variance is high, then highly connected nodes play an important role in determining which changes are propagated and which are blocked. If high-degree nodes are especially likely to be directly linked to other high degree nodes, as seems to be the case in social networks (but not other kinds of networks; Newman and Park 2003), they may in effect form a cluster of their own, enhancing their distinctiveness from low-degree nodes[28], and contributing to their social power; we expand on this below.

It is important to be clear that these features of network topology do not at all discriminate between different actions - while they make it more or less likely that actions *in general* will diffuse across a network, they do not make some actions more, and some actions less likely to diffuse. To properly understand how some actions are selected and others are not, one needs to look at differences in the *susceptibility of nodes to infection* by specific actions. If different actions are associated with differences in the resistance of nodes to contagion, than network effects will discriminate against some actions, and in favor of others.

[26] In fact, the most widely used methods for finding communities or clusters work by finding the borders across which diffusion is inhibited: Newman, 2006; Lee and Wasserman, 2010.

[27] All known social networks have some transitivity, but qualitative evidence suggests that there is considerable variation in how much clustering there is across time, space, and social classes. (See, e.g., Eagle et al. (2010) for some evidence that poverty and low social status are correlated with lack of diversity in networks in present-day England.) Many historical societies seem to have had much higher levels of transitivity than is typical of modern ones, with dense local communities that had few links between them. (This history of near-isolation is reflected in the continued genetic clustering revealed by genetic mapping (Cavalli-Sforza et al., 1994) and statistical studies such as Steffens et al. (2006).) Some modern societies still appear to have high degrees of clustering — for example, the Italian economy is less a coherent whole than a congeries of discrete local microeconomies, each with its own order and set of internal networked relations (Locke, 1995; Farrell, 2009). In such societies, we may expect deviant actions to spread locally with relative ease (if they pass the tests discussed above), but to be much less likely to be propagated at the national level.

[28] Cf. Gellner (1983, 1988) on the maintenance of cultural distinctions between ruling and ruled strata in complex pre-modern societies.

Building on our earlier arguments, the most plausible way to model this selection effect is to treat the resistance of nodes to infection as a function (perhaps with some additional random perturbation) of actors' preferences and bargaining strength. The intuition here is that the likelihood that an action will spread will depend on the desires of powerful actors (whether they favor it, or are against it) and their power to achieve their desires. Here, each interaction in which an actor seeks to employ a new action is treated as a node. The likelihood that the action will be successful - i.e. it will provide the actor employing it with a better payoff than she would otherwise have received - will depend both on the preferences of the other actor (whether she opposes the action, because it leads to lower payoffs than at the point associated with the usually expected action), and the power of the other actor to impose her preferred outcome by threatening or punishing to threaten. The outcome of this interaction can be observed by other actors associated with nodes that are directly linked to the original node. These actors will then either play out the same interaction (if the new action has been successful, encouraging the cognate actor at the new node to replicate it) or fail to play it out (if the new action has failed).

Preferences and bargaining strength (as determinants of node resistance), — combined with the features of the network topology that maps links across these nodes — provide us with an understanding of the selection effects that work to produce social institutions. As in the discussions of Knight (1992); Bowles and Naidu (2008) actions that are in the interests of powerful actors will be far more likely to be reproduced than those that run against these interests. However, specific attention to the *interactions* between network topology and resistance provide us with a better understanding of how the structure of communication may interact with power relations to aid or impede the dissemination of new actions, leading in turn to institutional change. It also points to questions that are barely addressed at all in the existing literature. For example, it allows us better to understand the relative speed of institutional change across a given community. Some network topologies will be more likely to facilitate swift diffusion of actions, and hence institutional change, than others. It also allows us better to understand how network structure (and in particular the level of transitivity) may lead to only partial institutional adoption, in which some clusters adopt a particular change and others do not.

We can go further. As discussed above, nodes with high degree are likely to have an unusually important role in determining which actions propagate and which do not. If these nodes differ in, e.g., their level of resistance, perhaps because they embody especially pronounced power asymmetries, then we may expect that change that disadvantages powerful actors is even less likely to occur. This may capture the important role that key nodes in communication networks play in determining which changes are propagated and which are blocked. Actors at these nodes hence have extraordinary political power, which may lead them to ally with, and obtain particular support from, other actors with the ability credibly to threaten violence or other unpleasant consequences for deviant actions. Gellner (1988) makes this point forcefully in his discussion of relations between specialists in violence and specialists in knowledge and communication in traditional societies. Sometimes, the latter may play a more important role than the former, for example in choosing new leaders,

because of their ability selectively to determine which possible behaviors are communicated, and which are not.[29]

Mathematically, if nodes of high degree belong disproportionately to a particular class which would be dis-advantaged by a new action, and so resist it, this has important consequences for the diffusion of the new action. In effect, the new action is trying to spread over a network from which the high-degree nodes have been preferentially removed (Albert et al., 2000; Callaway et al., 2000). Such "targeted" removal lowers the average degree of nodes on the effective network, as well as lowering the variance of degree; both of these make it harder for the innovation to spread, because (as we saw above) they raise the epidemic threshold. Preferentially blocking the high-degree nodes also changes the effective network for the diffusing innovation by increasing the average distance between nodes, especially between nodes which belong to highly transitive clusters. This is because the high-degree nodes are unusually apt to be long-range bridging ties (at least in many topologies), and in fact removing a fairly small fraction of the highest-degree nodes can fragment the network into disconnected components. The implication of all this is when an innovation disfavors a group of actors of unusually high degree, even without collective action on their part, their individual resistance to it creates disproportionate obstacles to the innovation's transmission, potentially blocking it completely.

This is not to say that these network effects (in conjunction with the mechanisms of variation discussed in the previous section) provide us with a simple and fully determinative model of institutional change. For one, the interactions between the different features of network topology that we discuss above have complicated and non-linear effects on the speed of diffusion. In networks with high transitivity, the addition of even a relatively very small number of random links can mean that diffusion is far quicker and more likely than before. Furthermore, even with these considerations, the diffusion of specific behaviors will remain not just stochastic, but subject to what can only be called historical contingency.[30] Both experiments (Salganik et al., 2006; Salganik and Watts, 2008) and models (Chamley, 2004) suggest that minor perturbations or differences in the early history of diffusion can have quite substantial consequences. Hence, while network structures may play an important general role in shaping patterns of diffusion and making broad classes of action more likely (e.g. actions that are favored by those with high bargaining power) or less likely than others, it will be difficult to predict whether any *specific* deviant action diffuses or not with any great certainty.

[29] Russell Hardin (1990) makes a similar argument in more modern contexts, when he points to the importance of securing control of means of mass communication (television and radio stations; the printing press) in revolutionary situations — whichever actor is able to control these nodes is likely to have a crucial advantage in shaping mass behavior. It will be interesting to see if this remains true in the face of communication technologies with more points of input, like cell phones or the Internet.

[30] The classic work on stochastic modeling of historically contingent processes is of course that of Brian Arthur and co-authors (Arthur, 1994); see Pemantle (2007) for a mathematically rigorous treatment of such models, Page (2006) for a conceptual discussion, and Walker (2007) for some of the statistical issues.

Innovation, Learning and Epistemic Democracy in Classical Athens

An influential tradition in political theory focuses on the benefits and shortcomings of 'epistemic democracy' - that is the of procedures of democratic decision making that are intended to elicit and aggregate useful knowledge from the population. As described by Joshua Cohen in an influential early piece,

> An epistemic interpretation of voting has three main elements: (1) an *independent standard* of correct decisions-that is, an account of justice or of the common good that is independent of current consensus and the outcomes of votes; (2) a *cognitive* account of voting - that is, the view that voting expresses beliefs about what the correct policies are according to the independent standard, not personal preferences for policies; and (3) an account of *decision making* as a process of the adjustment of beliefs, adjustments that are undertaken in part in light of the evidence about the correct answer that is provided by the beliefs of others. Thus, the epistemic conception treats processes of decision making as, potentially, rational processes of the formation of common judgments. (Cohen, 1986, p.34)

While work on the epistemic consequences of voting procedures, order of choices and so on continues (Dryzek and List, 2003; List, 2006), scholars of epistemic democracy have moved beyond voting, to consider the public benefits of deliberation and dissent. Anderson (2007) argues for the benefits of a Deweyan conception of democracy over both Condorcet-type voting models and Scott Page's 'diversity trumps expertise' theorem. Most recently, in a major new book, Ober (2008) uses a study of Athenian democracy in order to argue that democracies are better than other political orders at aggregating and distributing the knowledge held by their citizens. Ober documents how which Athenian political institutions served to gather useful knowledge that would otherwise have been unavailable to decision makers.

In this section, we take Ober's major claims, and recompose them using the various mechanisms laid out in the previous section. This is made considerably easier by the fact that we agree with Ober on many important questions - e.g. the relevance of network structure to explaining the diffusion of knowledge, the crucial importance of gathering dispersed knowledge *und so weiter*. However, where we disagree is on the role of evolutionary theory (which Ober invokes to explain Athens' relative success) and on the dynamics of institutional change. Specifically, we argue that the important evolutionary forces will apply at the level of the individual action, rather than the city (as a bundle of institutionalized responses to a particular strategic environment), and that the problems which knowledge-aggregation mechanisms must solve are dynamic rather than static. In other words, problem solving rather than merely reducing transaction costs, and hence bringing together existing dispersed information, should also be able to come up with new ideas and new information, hence allowing actors to arrive at new courses of action which may perhaps lead to

institutional change. [31]

The major goal of Ober (2008) is to explain the relationship between democratic procedures and *knowledge*, and hence to show that democracy has a consistent advantage over more hierarchical forms of political organization through promoting innovation and learning.In order to make this case, he makes two linked arguments. First, he seeks to show that Athens — the ancient world's most prominent democracy — was able to compete successfully against its less democratic rivals. Second, he seeks to show that Athens' specific advantages in this competition were linked to its democratic procedures of decision making.

To make the first argument hold, Ober invokes arguments about external competitive pressures. Political communities, like Athens, exist in 'multicommunity ecologies' which leads to competition among them. Borrowing from realist scholarship in international relations, he argues that the crucial level of selection is the level of inter-state competition. Ober finds that close to 12% of Greek city-states are known to have been destroyed at some point during the later archaic and classical periods, and problems of selection bias suggest that Greek city-states may have had as much as a one in three chance of suffering destruction at some point during this period. This competition did not only involve inter-state violence, but the possibility of such violence greatly heightened the stakes for decision makers within states. The "endemic threats to state existence from external rivals and internal civil war" led to "wide-spread agreement among the Athenians on the answer to the high-level question of "must our state seek to be relatively more prosperous, secure and powerful in relation to its rivals."" (Ober, 2008, p.100). More broadly, (Ober, 2006) claims that

> Ancient Greek city-states ... existed in a highly competitive environment in which failure was severely punished, by loss of independence or even annihilation. Destruction, total or partial, of physical infrastructure (sacking) or population (mass expulsion, extermination, or enslavement) was quite common: between a quarter and a third of the better-documented Greek states are known to have suffered such destruction at some point in their history. Poleis responded to such internal and external threats by experimenting with a variety of constitutional forms, with more and less extensive participation by citizens. Women never held active political citizenship in classical Greece, although they played very prominent roles in religion and other central aspects of civic life. But the extent of adult male participation varied widely, from the periodic experience of autocracy in Syracuse, to the more or less narrowly framed oligarchy in Corinth, to Athenss very robust and highly participatory democracy. These were essentially controlled political experiments, on a surprisingly large scale.

Hence, the "driver [of the day to day workings of democracy] was ... the imperative of deciding which knowledge and whose knowledge was most likely to ensure a common good (the survival and

[31] Our revised formulation also has the side-benefit, *contra* Anderson (2007), of showing how theoretical results about diversity can readily be incorporated in a dynamic Deweyan framework.

flourishing of the polis qua community) that was recognized by most as an essential prerequisite for the free pursuit of private interests by each individual" (Ober, 2008, p.102). This in turn led Athenians to create "effective strategies" to deal with problems of dispersed knowledge, unaligned actions and transaction costs.

We concentrate on Ober's account of the problem of dispersed knowledge, which is crucial to his intellectual project.[32] Here, Ober argues that there need to be proper incentives to share information, low communication costs to make it easier for useful information to spread, and an "epistemic sorting device" to distinguish useful from non-useful information.

Just as we do, Ober points to the importance of social networks in determining how information is shared. Athens originally consisted of a congeries of demes or village communities, within which individuals had strong social ties, and a high degree of mutual knowledge. After the reforms of Cleisthenes, however, these demes were joined together through constitutional arrangements that (perhaps inadvertently) created numerous bridging weak ties between these densely interconnected clusters. Artificial 'tribes' brought together demes with differing social characteristics, creating bonds among them through a variety of social rituals which promoted intermixing and the dissemination of knowledge. Members of the demes within a broader tribe were chosen by lot to serve on the advisory Council of 500 for a one year term, along with representatives of the other ten tribes within the city. Ober argues that these arrangements created ties both among the members of specific tribes, and among those serving together on the Council of 500 for a year. This led to the diffusion of knowledge, allowing it to "access, at fairly low cost, a good deal of the total knowledge available to the extended Athenian community." (Ober, 2008, p.101). The Council of 500, along with other social institutions (the seven hundred magistrates who served on boards for a year) created teams that could engage each other in intense ongoing interaction, while perhaps maximizing the likelihood that unique information could be presented, heard and incorporated. Other institutions, such as ostracism provided a potential check on the emergence of harmful leaders. Thus, Athenian institutions made it easier to gather information from a dispersed population, provided incentives to make citizens more knowledgeable and to share their knowledge as they participated in public office, and produced better policy than alternative, more hierarchical arrangements would have.

This account seems to provide a plausible understanding of how, in discrete and specific incidents of decision-making, institutions and networks may work together to aggregate and make useful information that would otherwise have remained inchoate. Such an account, however, falls short of Ober's stated ambition, which is to provide a theory of the relationship between institutional change and knowledge. While disclaiming any attempt at "[solving] the question of how much

[32]We believe that this is the most important contribution of the book. Ober's subsequent arguments about the need to ensure that actions are coordinated with each other, and that they minimize transaction costs are less novel, applying the claims of an extensive existing literature on the importance of focal points in generating common knowledge and relatively efficient institutions in lowering transaction costs. We do note, in passing that our arguments highlight the extreme unlikelihood of any very substantial form of common knowledge being generated across a society. Ober's major arguments on coordination do not in fact require common knowledge - simple shared expectations will be enough to ensure coordination in most of the situations he describes.

of the organizational design evidenced by the Athenian system arose from a series of adaptive experiments, and how much can be attributed to an intelligently chosen top-down blueprint," Ober still aspires to "put [the factors of knowledge] into a dynamic relationship with each other — to bring the system to life by explaining its endogenous capacity for change" (Ober, 2008, p.19).

However, in our judgment, Ober's account does not meet this (admittedly quite difficult) challenge. It never provides a systematic account of the forces underlying change in the Athenian system of decision-making, and why it was that Athenian democracy was consistently better than other political systems in ensuring not Athens not only captured static benefits, but also the dynamic benefits of ensuring that the process of information aggregation fed into endogenous processes of institutional change in mostly useful rather than mostly pernicious ways. Ober relies extensively on rational choice accounts of institutions (North et al., 2009; Greif, 2006), and we suspect that his account has difficulty in taking account of dynamics in large part because it leans on a theoretical framework in which "one is either in an equilibrium or undergoing change from one to another the way teleportation works in science fiction stories."[33]

There are surely elements of a more dynamic account in the book, but they do not accumulate to a satisfying story. Much of the time, Ober seems to be suggesting that the evolutionary forces driving selection within the state system conduct towards the selection of institutions within Athens (and perhaps other cities) that maximize the chances of successful flourishing. Thus, for example, he claims that '[c]ompetitive pressure rewards strong forms of state organization and drives out weak ones.' (Ober, 2008, p.13). However, here the problems are twofold. If the advantages of Athenian democracy as a superior form of organization were so obvious, why were other states with different forms either driven into extinction, or forced to emulate Athens' democratic institutions wholesale?[34] Ober's argument is that the 'full package of Athenian democracy in the Athenian style was not easy for other states to mimic because complex bodies of tacit knowledge do not travel well'(Ober, 2008, p.165), but, as Ober himself acknowledges, the importance of untransferable tacit knowledge is a ''hypothesis to be tested" (ibid) rather than a very strong basis for a general empirical claim. Nor did the imperative of city survival necessarily have an overwhelming effect on the microstructures of Athenian decision-making. As Ste. Croix notes, patterns of oligarchic-democratic politics often trumped city-state politics; "Oligarchic and democratic leaders had no hesitation in calling in outside powers to help them gain the upper hand against their adversaries," so that, for example, Corinthian democrats preferred to remain part of Argos than to regain independence on the condition of becoming an oligarchy.

[33]See Dorman (2008). This generalization is rather unfair to Greif (2006), which, building on Greif and Laitin (2004), provides an account of how equilibria can be self-reinforcing or self-undermining over time, depending on their effect on 'quasi-parameters.' But, Greif's evident creative intelligence notwithstanding, this account does as much to illustrate the limits of game theoretic accounts of institutional change as it does to expand those limits. More particularly, while Greif can tell us how equilibria corrode or reinforce themselves, he cannot tell us very much at all (Greif, 2006) about what happens *when* an equilibrium disappears, and what equilibrium may arise in its place. The next version of this paper will have a specific section dealing with these debates.

[34]Ober does suggest that Athens' edge over other cities was somewhat blunted as these cities emulated some of Athens' forms, but the extent of diffusion seems to have been relatively limited.

More seriously, there is good reason to believe that the selective pressures of the Greek interstate system were incapable of exerting the kind of micro-level selective impact that Ober posits. If we take the high end of Ober's estimate of the likelihood of city-state death - that a city-state has a 33% chance of being destroyed over a 500 year period, we can calculate the information provided by this selection mechanism using the procedure outlined in the paper's second section.[35] Ober's numbers work out to an annual risk of polis death of about 0.2 percent. This is not nothing, but it is not a very enormous risk either. If we applied selection like this to a population starting with a standard Gaussian distribution, removing the lowest 0.2 percent of the population each year, by the end of 500 years we would have increased the mean by about a standard deviation. Said another way, the entropy rate of selection is about 0.02 bits/year, so, accumulated over the whole of the archaic and classical periods, we get about 11 bits of information on which constellation of institutions works best. This is the *maximum* amount of information which could be extracted from the "signal" of city destruction (in real world situations we would plausibly expect rather less). Even if one is extremely optimistic, this paucity of information is almost certainly insufficient to do the job that Ober would like it to do.

None of this, is at all to say that Ober's ambition of understanding the dynamic consequences of the aggregation of knowledge is misguided or impossible. Far from it. If we retain Ober's goal, but shift our focus to a different level of selection, we can advance arguments which satisfy these information constraints in a much more plausible way. While selection is surely vital to the development of institutions, the selection which matters is not the rare destruction (or duplication) of whole polities or societies, but rather the continual choice by individuals about whether to conform to institutionalized expectations about their behavior, or to do something else. Renan famously claimed that a nation is a daily plebiscite; in a real sense, this is true of every human institution. They survive if, and only if, they are continually reproduced in human practices. This is where selection gets its purchase on institutions, just as we argue in the previous sections of this article. Accordingly, in the remainder of this section, we advance a sympathetic reconstruction of Ober's claims and evidence (including evidence from his other research on the topic) within this framework, so as better to capture the dynamics of change in Athenian society. Not only can this allow us to understand the dynamics of endogenous change, but it can help us to understand much more precisely how the initial organizational blueprint of the Athenian system (insofar as it shaped both network relations and the veto-points within these networks) intersected with a process of continuous adaptive experimentation to lead to change.

First — using the framework advanced in the earlier part of the article — we examine the mechanism of variation (in our account, the degree of heterogeneity in schemas possessed by actors in a given society). Ober provides substantial evidence suggesting that (a) there was relatively high diversity in the kinds of knowledge and social orientations possessed by various Athenian citizens, and (b) that there was significant and frequent contact between individuals who likely possessed different schemas. As Ober notes, the original Athenian demes were both very different

[35] The numbers here are rough: with a copy of Ober's original dataset we could likely do a better job.

from each other (some primarily agricultural, some located on the coastline, and more engaged with trade, some highly urbanized). The artificial 'tribe' system that was imposed by Cleisthenes on pre-existing social arrangementssystematically brought men from very different demes into close and sustained contact with each other.

> As a result, people with very different life histories and different sets of social and technical knowledge frequently found themselves in close social proximity to people they never would have otherwise known. The system very literally "intermixed" Athenians from different geographic/economic zones in a variety of psychologically powerful activities.(Ober, 2008, p.142)

While the evidence is understandably sparse, it appears to support Ober's broad conclusions.[36]

This diversity was plausibly enhanced in later decades by Athens' position as the central node of a large trading empire.[37] While it is difficult to quantify, there is a strong literature, dating back at least to John Stuart Mill, which argues that much of trade's salutary impact on the economy may be found in its tendency to expose people to different ideas and different ways of doing things.[38] Without a much more systematic comparison of Athens with other members of the Greek city-state system, it is impossible to say whether ϵ was higher in Athens than in its peers. But at the least, the evidence would seem to indicate that Athens' epsilon was not unusually low during the period of Athens' history that Ober describes (as he mentions, Athenians with at least some degree of partiality prided themselves on their unique talent for innovation). We may expect that variant actions were generated at a reasonably high rate, likely generating significant social frictions, but also sometimes providing pathways to points on the landscape of possibilities that would likely otherwise never have been observed.

Second, we turn to the mechanism of selection. Which structures selected over the variant actions that were propagated further, leading to institutional change, and which kind of actions were they likely to select? As discussed in previous sections, we argue that both (a) the kind of network structure and (b) the degree of asymmetry in power relations among particular actors embedded inside those structures will interact so as in general to select for certain kinds of actions, and against other kinds.

Happily for our purposes, Ober reconstructs the likely topology of the network shaping communication flows among Athenians after Cleisthenes' constitutional revolution. Although he does not

[36]For example, he points to evidence showing high rates of inter-deme marriage, suggesting strongly that social intercourse was not confined to specific demes. While it is risky to extrapolate social networks from historically scanty evidence, this would appear to contrast with the general picture indicated by genetic evidence of an extremely strong localism to patterns of sexual reproduction in pre-modern societies. It is very likely that Athens was highly unusual in the ancient world in the extent of social interactions across the city as a whole

[37]It was also arguably weakened by Pericles' reforms to limit citizenship to those with both Athenian fathers and mothers, rather than merely fathers as had previously been the case, and by other restrictions on citizenship

[38] See Mill (1998), "the economical advantages of commerce are surpassed in importance by those of its effects which are intellectual and moral. It is hardly possible to overrate the value, in the present low state of human improvement, of placing human beings in contact with persons dissimilar to themselves, and with modes of thought and action unlike those with which they are familiar. Commerce is now what war once was, the principal source of this contact."

specifically identify it as such, the broad network that he identifies is evidently a form of the small world topology described in Watts (1999) small-world networks, in which information can flow relatively rapidly and easily across sparse ties between densely interconnected clusters. Each deme (as a tightly connected community of actors who frequently interacted with each other) was a dense cluster within in the network. Ober identifies the ways in which tribal structures provided weak ties between specific subsets of demes (those in a particular tribe), and the Council of 500 (and perhaps the larger and more poorly organized Assembly, as well as juries, which had a significant political role) provide weak ties across the demes as a whole. Obviously, Ober's list is not exhaustive of network ties among Athenians - we may expect that economic and other social relations created weak ties across different demes. Nonetheless, Ober is surely right in arguing that these largely political networks played an important role in facilitating intellectual interchange about how e.g. to handle specific or recurring political problems. The weak political ties across demes and tribes thus likely helped disseminate more informal changes as well as structuring the more formal processes of institutional debate that Ober describes.

Most interesting for us, Ober provides evidence that Cleisthene's reforms and their aftermath helped to substantially dampen down asymmetries of bargaining power among Athenian citizens. Our arguments draw attention to one very important feature of Athenian democracy that Ober discusses, but that we believe likely to have played a more crucial role in shaping processes of institutional evolution than Ober himself recognizes.

As we have already noted, each deme chose its representatives to the Council by lot (i.e. through a random process of selection) for one year's service. If Ober is correct in arguing that these representatives played a key intermediating role in aggregating knowledge and in reshaping institutions, then this likely led to a highly unusual network configuration. Most non-modern societies (and, to a striking extent, modern societies too) are characterized by an effective, if somewhat uneasy alliance between the 'sword' and the 'book' (Gellner, 1988) - those who have an unusual degree of social power because of their ability to threaten violence (barons, princes, warlords and the like), and those who have an unusual degree of social power because of their control over information flows (clerisies in societies without mass literacy). Furthermore, as Newman and Park (2003) discuss, all known social networks are characterized by assortative mixing, under which nodes with high degree tend disproportionately to be associated with other nodes of high degree. More plainly put: those who are well connected themselves tend disproportionately to be connected to others who are well connected (this is partly, but not entirely, explained by group dynamics). If network structures and power relations are correlated so that (a) well-connected information brokers are well-connected to each other (and serve as the primary bridges between otherwise largely isolated clusters) and (b) so that these well-connected information brokers are also effectively allied with those who have physical force at their disposal, then we may expect that institutional innovations which are *not* in the interests of the elite are likely to be selected against, while those in the interests of the elite (and, perhaps, not in the interests of other actors in society) are likely to be selected for.

Athenian democracy, likely initially by accident, appears to have created institutions and network structures which greatly mitigated the power of elite citizens vis-a-vis other citizens.[39] By choosing major information brokers (members of the Council of 500) by lot from the general population of citizens, and only allowing them short terms of office, the reforms to the Council created a social network in which the political ties between different clusters ran through individuals who were broadly representative of citizens as a whole, rather than a clerical elite or a professional political class. The results were salutary. Those who *were* arguably professional politicians of a kind were typically obliged to appeal to the citizenry rather than to aristocratic norms (Ober, 1991). More directly, our arguments provide a mechanism through which the spirit of innovation that Athenians laid claim to was plausibly linked in a quite direct way to Athenian democratic norms. Because traditional elites did not have a chokehold over the major points of communication and knowledge aggregation, they were not able as easily to block or stifle reforms that were in the general interests of Athenian citizens as their equivalents in other, less democratic cities. Such innovations were more likely to be pushed through, as long as they proved persuasive to citizens' representatives in the Council of 500, and to citizens as a general body when they met in the Assembly (where each citizen had in principle an equal right to be heard in debate). In short, our arguments about selection would suggest that Athens was more likely to embrace innovations that were of general social use than the typical agrarian societies described by modern sociologists such as Gellner and indeed by classical sources such as Xenophon, in which elites and court politics played a preponderant role in dictating the pace of social innovation. The random assignment of information brokerage thus perhaps allowed the city to take far greater advantage of a degree of internal heuristic diversity which was likely greater in any event than that of major rivals such as Sparta.[40]

Random assignment of Council membership was not the only way in which Athenian democracy shook up its internal social networks in heuristically useful ways. The provision of pay for Council service also helped encourage participation from a broad range of actors from different social circumstances[41] The relationship between citizenship and the deme also perhaps helped rebalance power relations. An applicant for citizenship would be presented to the deme's assembly, who would vote (on the basis of their knowledge of whether the applicant fulfilled the birth requirements) on whether to accept the applicant or not, creating "new ties [which] cut across the horizontal strata of birth and wealth," so that even the richest and most socially powerful began their political life when they were "voted into the demos by the decision of [their] ordinary neighbors" (Ober, 1991, p.70). Finally, even while some offices of state retained property qualifications (e.g. the archonship), they were chosen by lot from among a larger body of men (Ober, 1991). It is far from clear that these

[39] As democracy eventually started to decline, power asymmetries increased. In the description of Ste. Croix, "The most significant result of the destruction of Greek democracy was the complete disappearance of the limited measure of political protection allowed to the lower classes against exploitation by the propertied."

[40] It is perfectly possible that this resulted in a 'noisier' political system in the specific sense that the filters for institutional innovation were looser than under other possible political arrangements, leading to greater inconsistency over time. Whether or not such noisiness is worse than the alternative depends on whether some degree of randomness in outcome is better or worse than a repeated bias in decisions in favor of the interests of more powerful actors.

[41] Although if Aristophanes' exaggerated and polemic take on the jury system is not *entirely* misleading, the introduction of dikastic pay may have led to a narrowing rather than widening of the class of jurors.

social institutions were *intended* to rebalance power relations, and they were surely not designed so as to increase the likelihood that specific measures and laws would innovate in the interests of the citizenry as a whole. Even so, they plausibly had just this consequence.

In conclusion, we should clearly acknowledge that our arguments here indicate plausible mechanisms rather than demonstrating that these mechanisms in fact obtained. We largely rely on Ober's own evidence (as we note, while we disagree with some of the underlying mechanisms that he posits, we are happy to embrace many of the major points of his argument), the secondary literature, and the ruins of our own classical educations rather than original research. A proper investigation of these arguments would at a minimum require considerable comparative work to examine whether specific features such as lotteries for office were systematically associated with more generally benign institutions.[42] Even so, we believe that our framework not only provides a more solid grounding for Ober's project, but also points the way more clearly toward such comparative work. By elucidating more precisely the processes through which evolutionary forces can typically lead to the selection of some innovations over others, we show the way towards a much easier set of microcomparisons. Rather than seeking to compare political systems as a whole, our arguments highlight the particular aspects of political systems that influence potential choke-holds over information flows, and provides a means for assessing their likely effects. This seems to us to be a sounder basis for (as Adam Przeworski describes the comparative method) "testing hypotheses across different historical and geographical conditions."

While our major purpose in this paper is empirical, our findings have some normative implications that are not too difficult to draw out. While, like Ober, we recognize the great merits of Athenian democracy compared to most of its contemporaries, we share his abhorrence for institutions such as slavery and the subjection of women. As we repeatedly note above, our argument suggests only that any beneficial consequences of Athenian democracy applied primarily to Athenian citizens - since only their interests were taken into account in the social institutions that we have discussed (women could not even testify in their own defense in trials), we may expect that others' interests received relatively short shrift. To say that Athenian democratic institutions facilitated innovations that redounded to the advantage of Athenian citizens is not to mistake Athenian democracy for our own ideals (even in the very limited sense of our paper, which discusses political theory in the context of a quite limited theory of democracy and knowledge, rather than e.g. theories of justice).

Even so — and even under these normative constraints — we suspect that Ober's theory harbors wider implications than it might sometimes seem from Ober's own presentation. We doubt that he would disagree even slightly with the claim that the cognitive benefits of Athenian democracy were limited by its systematic exclusion of slaves, women and metics from political debate. The greater the diversity that can be processed by institutions of knowledge aggregation, the more likely it

[42] We regret that Eric W. Robinson's survey of democracy in antiquity was not publicly available in time for us to read it for this version of the paper.

is that otherwise unreachable points on the fitness landscape will be discovered.[43] Yet Ober also argues that the cognitive benefits of democratic decision processes can be captured within e.g. hierarchically organized firms operating in markets. We are somewhat skeptical of this claim — efforts to harness democratic methods of knowledge aggregation within businesses have in general been disappointing (Henwood, 2003). *Contra* Ober, the employees of a firm are *not* its citizens, because profits flow to, and ultimate legal control rests with, the owners, though in practice they usually delegate to a largely self-perpetuating management. To draw a more appropriate Athenian analogy, employees are metics, foreigners (or descendants of foreigners) who are allowed to live in the city, work there, serve in the military, pay taxes, and generally tie their fortunes to it, but are not even second-class citizens.

Our perhaps somewhat impertinent reconstruction of Ober - if taken seriously - leads in a more radical direction. If his arguments, as we have revised them, are right, they suggest that truly capturing the benefits of cognitive diversity, and aggregating the knowledge of a wide variety of different individuals with different schemas, requires quite radical democratic institutions, and a much greater equality of voice. They require, in short, a Deweyan democracy, combining radical, persistent experimentation together with a systematic willingness to retain the forms of diversity from which new ideas for social experiments arise, and a set of filtering mechanisms that are designed, insofar as is compatible with a bare minimum of order, to allow a great variety of perspectives to have their say.[44] Perhaps the most important lesson that Athenian democracy has to offer, on our synthesis, is its tentative demonstration of how particular measures that increase representation of diversity, most prominently a combination of office-by-lottery with pay, may serve to greatly further socially useful innovation. While other scholars have drawn attention to the possible benefits of random assignment for e.g representativeness and consequent democratic legitimacy, the possible benefits for decision making have been understudied. Studying the mechanisms through which innovations arise, and through which some innovations are selected and others fail, provides us with a very useful way to capture this important dimension of political institutions.

Institutional Change and Political Economy - Beyond Equilibrium Explanations

TO BE WRITTEN IN NEXT DRAFT

Conclusions

This paper proposes a new way of thinking about the dynamics of institutional change. Evolutionary theory, as it has been developed by biologists, points to mechanisms of transmissible variation and selection. Mathematical modeling shows how these mechanisms may be understood as a search

[43] We assume, for the sake of argument, that it is typically a *good thing* ceteris paribus when new points can be discovered - there are of course some points that we might prefer *ex post* to have been left undiscovered.

[44] We here borrow from and concur with Johnson and Knight, forthcoming.

for answers to problems posed by a selection process across a solution landscape whose topology is unknown. These models provide general lessons that can be applied outside evolutionary biology. In this paper, we provide one such application, in which cognitive heterogeneity provides variation, and bargaining relations, as it intersects with network topology, selects over these variations. We show how this application may help elucidate and extend further the findings of an important recent literature in political theory. The next version of this article will similarly seek to speak to the political economy of institutional development, showing how these arguments help explain change in ways that current, game-theoretic and transaction-cost based approaches cannot.

The intellectual advantages of this kind of approach, are, we hope, reasonably clear. In contrast to some currently dominant approaches, the dynamics of change are built directly into the explanatory framework. There may be relatively stable 'states' in our account, which bear an initial resemblance e.g. to game theoretic equilibria. However, they do not have the same self-reinforcing characteristics as equilibria, in which actors' expectations and strategies converge so that it is nearly impossible to see how they might change, without some kind of external shock. In our argument, in contrast, stability is always provisional - actors never agree in their interpretation and strategy in the strong sense that they do in game theory, and always retain the capacity for unexpected action. While most of these actions are likely to fail, some may provide paths to new stable states and hence new institutions if they propagate across the population. Institutions are *always contingent* — even when they are not the subject of direct contention, their reproduction can never be taken for granted. We note that this is likely to be true not only for our arguments, but for a family of evolutionary explanations, some of which may look to quite different mechanisms (one of our aims in this paper is to invite others, who disagree with us e.g. on the importance of power asymmetries, or heterogeneity, or network topology, to construct alternative evolutionary explanations using the kinds of very broad concepts we explore in the paper's second section).

Here, we have much in common with historical institutionalists, who are directly concerned with questions of institutional reproduction (and indeed, with how institutions may change as they reproduce). However, our account tries to build foundations for a task that historical institutionalism does not aspire to — to systematically assess the relative degrees of openness of different societies to change, and to assess further which *kinds* of change, and in whose interest, are likely under different societal configurations. This cuts against current trends in historical institutionalism, which has come to focus more directly on assessing specific mechanisms of institutional change, and on distinguishing clearly between ideational (or in our language cognitive) and power-based accounts of change. While we believe that there is the potential for valuable dialogue, we are not currently entirely sure which direction it will go.

This brings us directly to the most obvious thin point in our account. It has little in the way of direct discussion of human agency. While agents play a key role in our argument (we are intensely interested in how individuals, and the differences between them, can have important political consequences), our major arguments implicate structural features of society (the degree of heterogeneity among agents, network topology, the respective bargaining strength of different

social classes of agents) rather than agents themselves. Put differently, our model is rather better at talking about the general circumstances under which agency is more or less likely to lead to institutional change than in talking about agency itself.

This means that we are likely to ask different questions and look for different kinds of answers than many other scholars interested in institutional change (it does not, of course, mean that we have nothing to say to these scholars, nor that they have nothing to say to us). Yet these questions are, it would seem to us, good ones. As the Ernest Gellner book that we cite at the beginning of this paper notes, there is precious little interest in "philosophic history" among sociologists, political scientists and indeed historians these days, and there is rather less today than there was when Gellner wrote those words. Large scale inquiries into the nature of social change have fallen out of fashion. Even so, what is unfashionable may still be important. And there are few more important questions than those concerning the large scale constraints on human social evolution. Even if our answers are still in many respects preliminary, they are surely answers to questions that are worth trying to answer.

References

Adami, C. and Cerf, N. J. (2000). Physical complexity of symbol sequences. *Physica D*, 137:62–9.

Adcock, R. (2009). Rethinking classical liberalism in "progressive" times: The divergent sociologies of spencer and sumner. Paper presented at the Annual Meeting of the American Political Science Association, Toronto, Canada.

Albert, R., Jeong, H., and Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406:378–382.

Anderson, E. (2007). The epistemology of democracy. *Episteme: A Journal of Social Epistemology*, 3(1):8–22.

Arthur, W. B. (1994). *Increasing Returns and Path Dependence in the Economy*. University of Michigan Press, Ann Arbor.

Badii, R. and Politi, A. (1997). *Complexity: Hierarchical Structures and Scaling in Physics*. Cambridge University Press, Cambridge, England.

Balkin, J. M. (1998). *Cultural Software: A Theory of Ideology*. Yale University Press, New Haven, Connecticut.

Bannister, R. C. (1989). *Social Darwinism: Science and Myth in Anglo-American Thought*. Temple University Press, Philadelphia.

Barry, B. (1988). *Sociologists, economists and democracy*. University of Chicago Press.

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press, Cambridge, England.

Bednar, J. and Page, S. (2007). Can game(s) theory explain culture? the emergence of cultural behavior within multiple games. *Rationality and Society*, 19:65–97.

Benaïm, M. and Weibull, J. W. (2003). Deterministic approximation of stochastic evolution in games. *Econometrica*, 71:879–903.

Bergstrom, C. T. and Lachmann, M. (2005). The fitness value of information. E-print, arxiv.org.

Bergstrom, C. T. and Rosvall, M. (2009). The transmission sense of information. *Biology and Philosophy*, online before print:1–18.

Bonner, J. T. (1988). *The Evolution of Complexity, by Means of Natural Selection*. Princeton University Press, Princeton, New Jersey.

Boudon, R. (1984/1986). *Theories of Social Change: A Critical Appraisal*. Polity Press, Cambridge, England. Translated by J. C. Whitehouse from *La place du désordre*, Paris: Presses Universitaires de France.

Bowles, S. (2004). *Microeconomics: Behavior, Institutions, and Evolution*. Princeton University Press, New York.

Bowles, S. and Gintis, H. (1998). The moral economy of communities: Structured populations and the evolution of prosocial norms. *Evolution and Human Behavior*, 19:3–25.

Bowles, S. and Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous groups. *Theoretical Population Biology*, 65:17–28.

Bowles, S. and Naidu, S. (2008). Persistent institutions. Technical Report 08-04-015, Santa Fe Institute.

Boyd, R. and Richerson, P. J. (1985). *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.

Boyd, R. and Richerson, P. J. (2005). *The Origin and Evolution of Cultures*. Oxford University Press, Oxford.

Burt, R. S. (1992). *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, Massachusetts.

Callaway, D. S., Newman, M. E. J., Strogatz, S. H., and Watts, D. J. (2000). Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, 85:5468–5471.

Cavalli-Sforza, L. L., Menozzi, P., and Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton University Press, Princeton.

Chamley, C. (2004). *Rational Herds: Economic Models of Social Learning*. Cambridge University Press, Cambridge, England.

Cohen, J. (1986). An epistemic conception of democracy. *Ethics*, 97(1):26–38.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley, New York.

Crutchfield, J. P. and Schuster, P., editors (2003). *Evolutionary Dynamics: Exploring the Interplay of Selection, Accient, Neutrality, and Function*, Oxford. Oxford University Press.

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press, Oxford.

Dawkins, R. (1982). *The Extended Phenotype*. Oxford University Press, Oxford. Re-issued with an afterword by Daniel Dennett and bibliographic supplement, 1999.

Dobbin, F. (1997). *Forging industrial policy: The United States, Britain, and France in the railway age*. Cambridge Univ Pr.

Dorman, P. (2008). What would a scientific economics look like. *Real-world economics review*, 47:166–172.

Draief, M. and Massoulié, L. (2010). *Epidemics and Rumors in Complex Networks*. Cambridge University Press, Cambridge, England.

Dryzek, J. S. and List, C. (2003). Social choice theory and deliberative democracy: a reconciliation. *British Journal of Political Science*, 33(01):1–28.

Dupuis, P. (1988). Large deviations analysis of some recursive algorithms with state-dependent noise. *Annals of Probability*, 16:1509–1536.

Durrett, R. (2006). *Random Graph Dynamics*. Cambridge University Press, Cambridge, England.

Eagle, N., Macy, M., and Claxton, R. (2010). Network diversity and economic development. *Science*, 328:1029–1031.

Easley, D. and Kleinberg, J. (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambidge University Press, Cambridge, England.

Eggertsson, T. (1990). *Economic Behavior and Institutions*. Cambridge University Press, Cambridge, England.

Elster, J. (1985). *Making Sense of Marx*. Cambridge University Press, Cambridge, England.

Engel, A. and Van den Broeck, C. (2001). *Statistical Mechanics of Learning*. Cambridge University Press, Cambridge, England.

Farrell, H. (2009). *The Political Economy of Trust: Institutions, Interests and Inter-Firm Cooperation in Italy and Germany*. Cambridge University Press, Cambridge, England.

Feldman, D. P. and Crutchfield, J. P. (1998). Measures of statistical complexity: Why? *Physics Letters A*, 238:244–252.

Fisher, R. A. (1958). *The Genetical Theory of Natural Selection*. Dover, New York, second edition. First edition published Oxford: Clarendon Press, 1930.

Fligstein, N. and McAdam, D. (2011). Toward a General Theory of Strategic Action Fields*. *Sociological Theory*, 29(1):1–26.

Freidlin, M. I. and Wentzell, A. D. (1998). *Random Perturbations of Dynamical Systems*. Springer-Verlag, Berlin, second edition. First edition first published as *Fluktuatsii v dinamicheskikh sistemakh pod deistviem malykh sluchainykh vozmushchenii*, Moscow: Nauka, 1979.

Gellner, E. (1983). *Nations and Nationalism*. Blackwell, Oxford.

Gellner, E. (1988). *Plough, Sword and Book: The Structure of Human History*. University of Chicago Press, Chicago.

Gelman, A. (2007). Letter to the editors regarding some papers of Dr. Satoshi Kanazawa. *Journal of Theoretical Biology*, 245:597–599.

Gershenson, C. and Lenaerts, T. (2008). Evolution of complexity. *Artificial Life*, 14:241–243. Introduction to special issue on the evolution of complexity.

Gould, S. J. (2002). *The Structure of Evolutionary Theory*. Harvard University Press, Cambridge, Massachusetts.

Greif, A. (1994). Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. *The Journal of Political Economy*, 102(5):912–950.

Greif, A. (2006). *Institutions and the path to the modern economy: Lessons from medieval trade*. Cambridge University Press, Cambridge, England.

Greif, A., Milgrom, P., and Weingast, B. R. (1994). Coordination, commitment, and enforcement: The case of the merchant guild. *The Journal of Political Economy*, 102(4):745–776.

Greif, A. D. and Laitin, D. D. (2004). A theory of endogenous institutional change. *American Political Science Review*, 98:633–652.

Hacker, J. S. (2004). Privatizing risk without privatizing the welfare state: The hidden politics of social policy retrenchment in the United States. *American Political Science Review*, 98(02):243–260.

Haldane, J. B. S. (1954). The measurement of natural selection. In *Proceedings of the 9th International Congress of Genetics*, volume 1, pages 480–487.

Hall, P. and Thelen, K. (2009). Institutional change in varieties of capitalism. *Socio-Economic Review*, 7(1):7.

Hardin, R. (1990). The social evolution of cooperation. In Cook, K. S. and Levi, M., editors, *The Limits of Rationality*. University of Chicago Press, Chicago.

Hart, H. L. A. (1994). *The concept of law*. Clarendon Press, Oxford.

Hayek, F. A. (1937). Economics and knowledge. *Economica*, 4:33–54. Reprinted in (Hayek, 1948, pp. 33–56).

Hayek, F. A. (1948). *Individualism and Economic Order*. University of Chicago Press, Chicago.

Henwood, D. (2003). *After the New Economy*. New Press.

Hodgson, G. and Knudsen, T. (2008). Information, complexity and generative replication. *Biology and Philosophy*, 23(1):47–65.

Hodgson, M. G. S. (1993). *Rethinking World History: Essays on Europe, Islam, and World History*. Cambridge University Press, Cambridge, England. Edited, with an introduction and conclusion, by Edmund Burke III.

Hull, D. L. (1988). *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. University of Chicago, Chicago.

Hume, D. (1994). *Political Essays*. Cambridge University Press, Cambridge, England. Edited by Knud Haakonssen.

James, W. (1880). Great men, great thoughts, and the environment. *The Atlantic Monthly*, 46 (276):441–459. Reprinted as "Great Men and Their Environment" in *The Will to Believe and Other Essays in Popular Philosophy*, New York: Longmans, Green and Co., 1897.

Johnson, J. (2002). How conceptual problems migrate: Rational choice, interpretation, and the hazards of pluralism. *Annual Review of Political Science*, 5:223–248.

Kenah, E. and Robins, J. M. (2007). Second look at the spread of epidemics on networks. *Physical Review E*, 76:036113.

Knight, J. (1992). *Institutions and Social Conflict*. Cambridge University Press, Cambridge, England.

Knight, J. and North, D. (1997). Explaining Economic Change:¡ i¿ The Interplay Between Cognition and Institutions¡/i¿. *Legal Theory*, 3(03):211–226.

Kussell, E. and Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309:2075–2078.

Lee, A. B. and Wasserman, L. (2010). Spectral connectivity analysis. *Journal of the American Statistical Association*, 105:1241–1255.

Lewis, D. K. (2002). *Convention: A philosophical study.* Wiley-Blackwell.

List, C. (2006). The discursive dilemma and public reason. *Ethics*, 116(2):362–402.

Locke, R. M. (1995). *Remaking the Italian Economy.* MIT Press, Cambridge, Massachusetts.

Lotka, A. J. (1924). *Elements of Physical Biology.* Williams and Wilkins, Baltimore, Maryland. Reprinted as *Elements of Mathematical Biology*, New York: Dover Books, 1956.

Lynch, A. (1996). *Thought Contagion: How Belief Spreads Through Society.* Basic Books, New York.

Mahoney, J. and Thelen, K. (2009). *Explaining institutional change: ambiguity, agency, and power.* Cambridge University Press.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444.

McShea, D. W. (1996). Metazoan complexity and evolution: Is there a trend? *Evolution*, 50:641–668.

McShea, D. W. and Brandon, R. N. (2010). *Biology's First Law: The Tendency for Diversity and Complexity to Increase in Evolutionary Systems.* University of Chicago Press, Chicago.

Milgrom, P. R., North, D. C., and Weingast, B. (1990). The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Economics & Politics*, 2(1):1–23.

Mill, J. S. (1998). *Principles of political economy.* Oxford University Press, USA.

Miller, A. S. and Kanazawa, S. (2007). *Why Beautiful People Have More Daughters.* Perigee, New York.

Morgan, S. L. and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research.* Cambridge University Press, Cambridge, England.

Nelson, R. R. and Winter, S. G. (1982). *An Evolutionary Theory of Economic Change.* Harvard University Press, Cambridge, Massachusetts.

Newman, M. E. J. (2002). The spread of epidemic disease on networks. *Physical Review E*, 66:016128.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45:167–256.

Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74:036104.

Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press, Oxford, England.

Newman, M. E. J. and Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E*, 68:036122.

Nilsson, M. and Snoad, N. (2000). Error thresholds for quasispecies on dynamic fitness landscapes. *Physical Review Letters*, 84:191–194.

Nilsson, M. and Snoad, N. (2002a). Optimal mutation rates in dynamic environments. *Bulletin of Mathematical Biology*, 64:1033–1043.

Nilsson, M. and Snoad, N. (2002b). Quasispecies evolution on a fitness landscape with a fluctuating peak. *Physical Review E*, 65.

Niyogi, P. (2006). *The Computational Nature of Language Learning and Evolution*. MIT Press, Cambridge, Massachusetts.

North, D., Wallis, J., Weingast, B., and dawsonera (2009). *Violence and social orders: a conceptual framework for interpreting recorded human history*. Cambridge University Press.

North, D. C. and Thomas, R. P. (1973). *The Rise of the Western World: A New Economic History*. Cambridge University Press, Cambridge, England.

Ober, J. (1991). *Mass and elite in democratic Athens: Rhetoric, ideology, and the power of the people*. Princeton University Press.

Ober, J. (2006). Learning from Athens: Democracy by design. *Boston Review*, 31:13–17.

Ober, J. (2008). *Democracy and knowledge: innovation and learning in classical Athens*. Princeton University Press.

Ostrom, E. and Basurto, X. (2009). The Evolution of Institutions: Towards a New Methodology.

Page, S. E. (2006). Path dependence. *Quarterly Journal of Political Science*, 1:87–115.

Page, S. E. (2007). *The Difference: How the Power of Diveristy Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press, Princeton, New Jersey.

Page, S. E. (2008). Uncertainty, difficulty, and complexity. *Journal of Theoretical Politics*, 20:115–149.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146.

Pemantle, R. (2007). A survey of random processes with reinforcement. *Probability Surveys*, 4:1–79.

Pepper, J. W. and Smuts, B. B. (2000). The evolution of cooperation in an ecological context: an agent-based model. In Kohler, T. A. and Gumerman, G. J., editors, *Dynamics in Human and Primate Societies: Agent-Based Modeling of Social and Spatial Processes*, pages 45–76, Oxford. Oxford University Press.

Pepper, J. W. and Smuts, B. B. (2002). A mechanism for the evolution of altruism among non-kin: positive assortment through environmental feedback. *American Naturalist*, 160:205–212.

Popper, K. R. (1960). *The Poverty of Historicism*. Routledge, London, 2nd edition. First edition, 1957.

Quiñonero-Candela, J., Sugiyama, M., Schwaighofer, A., and Lawrence, N. D., editors (2009). *Dataset Shift in Machine Learning*. MIT Press, Cambridge, Massachusetts.

Richerson, P. J. and Boyd, R. (2006). *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press, Chicago.

Rivoire, O. and Leibler, S. (2010). The value of information for populations in varying environments. Electronic pre-print, arxiv.org.

Rubinstein, A. (1991). Comments on the interpretation of game theory. *Econometrica*, 59:909–924.

Runciman, W. G. (1983–1997). *A Treatise on Social Theory*. Cambridge University Press, Cambridge, England. In three volumes.

Runciman, W. G. (1986). On the tendency of human societies to form varieties. *Proceedings of the British Academy*, 72:149–165.

Runciman, W. G. (1998). *The Social Animal*. HarperCollins, London.

Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311:854–856.

Salganik, M. J. and Watts, D. J. (2008). Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Social Psychological Quarterly*, 71:338–355.

Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, Princeton.

Sawai, S., Thomason, P. A., and Cox, E. C. (2005). An autoregulatory circuit for long-range self-organization in *dictyostelium* cell populations. *Nature*, 433:323–326.

Schacter, D. L., Norman, K. A., and Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, 49:289–318.

Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard University Press, Cambridge, Massachusetts.

Schiemann, J. W. (2000). Meeting halfway between Rochester and Frankfurt: generative salience, focal points, and strategic interaction. *American Journal of Political Science*, 44:1–16.

Shalizi, C. R. (2009). Dynamics of Bayesian updating with dependent data and misspecified models. *Electronic Journal of Statistics*, 3:1039–1074.

Shalizi, C. R. and Crutchfield, J. P. (2001). Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, 104:817–879.

Shalizi, C. R., Klinkner, K. L., and Haslinger, R. (2004). Quantifying self-organization with optimal predictors. *Physical Review Letters*, 93:118701.

Shalizi, C. R. and Thomas, A. C. (2010). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research*, forthcoming.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423. Reprinted in Shannon and Weaver (1963).

Shannon, C. E. and Weaver, W. (1963). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, Illinois.

Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach*. Basil Blackwell, Oxford.

Spruyt, H. (1996). *The sovereign state and its competitors: an analysis of systems change*. Princeton University Press.

Steffens, M., Lamina, C., Illig, T., Bettecken, T., Vogler, R., Entz, P., Suk, E.-K., Toliat, M. R., Klopp, N., Caliebe, A., König, I. R., Köhler, K., Lüdemann, J., Lacava, A. D., Fimmers, R., Lichtner, P., Ziegler, A., Wolf, A., Krawczak, M., Nürnberg, P., Hampe, J., Schreiber, S., Meitinger, T., Wichmann, H.-E., Roeder, K., Wienker, T. F., and Baur, M. P. (2006). SNP-based analysis of genetic substructure in the German population. *Human Heredity*, 62:20–29.

Thelen, K. A. (2004). *How institutions evolve: The political economy of skills in Germany, Britain, the United States, and Japan*. Cambridge Univ Pr.

Toulmin, S. (1972). *Human Understanding: The Collective Use and Evolution of Concepts*. Princeton University Press, Princeton, New Jersey.

Turner, S. P. (1994). *The Social Theory of Practices: Tradition, Tacit Knowledge, and Presuppositions*. University of Chicago Press, Chicago.

van Nimwegen, E. and Crutchfield, J. P. (2000). Metastable evolutionary dynamics: Crossing fitness barriers or escaping via neutral paths? *Bulletin of Mathematical Biology*, 62:799–848.

van Nimwegen, E., Crutchfield, J. P., and Mitchell, M. (1997). Finite populations induce metastability in evolutionary search. *Physics Letters A*, 229:144–150.

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. Springer-Verlag, Berlin, 2nd edition.

Vidyasagar, M. (2003). *Learning and Generalization: With Applications to Neural Networks*. Springer-Verlag, Berlin, second edition.

Walker, R. (2007). Path, phat, and state dependence in observation-driven Markov models. E-print, Society for Political Methodology archive.

Watkin, T. L. H., Rau, A., and Biehl, M. (1993). The statistical mechanics of learning a rule. *Reviews of Modern Physics*, 65:499–556.

Watts, D. J. (1999). *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, Princeton, New Jersey.

Watts, D. J. (2004). The "new" science of networks. *Annual Review of Sociology*, 30:243–270.

Wiesner, K. and Ladyman, J. (2010). What is a complex system? *European Journal for Philosophy of Science*, submitted.

Williams, G. C. (1966). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press, Princeton. Republished, with a new preface by the author, 1996.

Wilson, E. O. (1999). *Consilience: The Unity of Knowledge*. Vintage, New York.

Young, H. P. (1998). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton.