Facing Uncertainty:
Norms and Formal Institutions as Shared Mental Models

Presented at
SASE Mini-Conference on Uncertain Futures in Economic Decision Making
London School of Economics
2 July 2015

Revised 25 June 2019

William D. Ferguson
Gertrude B. Austin Professor of Economics
Grinnell College
ferguso1@grinnell.edu

Keywords: mental models, institutions, uncertainty, bounded rationality

ABSTRACT

This paper presents a theoretical argument focused on how social norms and formal institutions operate as cognitive coping mechanisms among groupings of boundedly rational actors who face fundamental uncertainty concerning their political and economic environments. Broadly speaking, informal and formal institutions facilitate strategic decision making by coordinating agents' understandings of their social environments and their conceptions of how myriad actions of involved participants (including themselves) may affect such environments, along with their positions and wellbeing. Yet institutions and the associated cognitive processes, are subject to periods of rapid transformation that sometimes exhibit properties of cascading imitation across individuals and groups.

After addressing background concepts, this paper makes four related assertions. First, heuristics and, by extension, mental models respond to shared narratives in a fashion that often generates conformity of belief and action. Second, mental models follow the dynamics of punctuated equilibrium processes. Third, institutions are a type of shared mental model that convey basic understandings across uncertain environments. Fourth, by enabling boundedly rational actors to manage uncertainty, institutions effectively choreograph social activity. Discussion includes reference to classical, evolutionary and epistemic game-theoretic modeling. Social choreography thus follows punctuated equilibrium dynamics that offer relative predictability during stable phases and stark uncertainty during rapid phases of punctuation. The paper closes with a fewon implications on the political economy of institutional change.

Introduction

Uncertainty has the potential to reduce strategic decision making among boundedly rational actors to mere uneducated guesswork, presenting daunting collective-action problems to groups of individuals who might hope to jointly cooperate in various endeavors. To navigate uncertainty, individuals use heuristics (conceptual rules of thumb) and mental models (formed conceptions of categories and causal relationships) that are often shared and transmitted across (and among) groups. Narratives, as stories about cause and effect, often shape such transmissions. On a broader scale, social norms and formal institutions emerge from and reflect the sharing of mental models. As such, they operate as cognitive coping mechanisms among groupings of boundedly rational actors that facilitate strategic decision making and effectively choreograph activity by shaping agents' understandings of their social environments, associated interactions, and possible outcomes. In so doing, norms and formal institutions reduce uncertainty to manageable levels that facilitate relative predictability and coordination. Yet occasional rapid bursts of change that sometimes exhibit properties of cascading imitation undermine predictability and coordination. Narratives play a role in all of these dynamics.

This paper makes four related assertions. First, heuristics and, by extension, mental models respond to shared narratives in a fashion that can generate conformity of belief and action. Second, mental models follow the dynamics of punctuated equilibrium processes. Third, institutions are a type of shared mental model that convey basic understandings across uncertain environments. Fourth, by enabling boundedly rational actors to manage uncertainty, institutions effectively choreograph social activity.

The paper uses variations of contemporary game-theoretic reasoning to illustrate and substantiate elements of these four assertions. Classical game theory (founded a concept of substantive rationality whereby individual agents possess sufficient cognitive ability to estimate best responses to strategic encounters) can illustrate conformity effects implied by the first assertion.[1] Multi-player games of assurance offer relatively simple illustrations. Evolutionary game-theoretic reasoning—in which agents "inherit" certain behavioral or cognitive practices (from prior acculturation) and then, over time, adapt to encountered circumstances—can represent learning properties that accompany the emergence of heuristics and mental models and their punctuated equilibrium proprieties. Epistemic games can represent how institutions can, via "signals," choreograph understandings and responses across groups of individuals.

The paper proceeds to link the four assertions together in a discussion of social choreography, imitation, conformity, and predictability, noting relations to punctuated-equilibrium dynamics. The conclusion offers a brief statement concerning the generation and (often partial) resolution of collective-action problems on both behavioral and cognitive levels.

Background Concepts

Subsequent discussion builds on the following concepts. *Uncertainty* implies that agents do not know underlying probability distributions that specify the likelihood of possible outcomes that arise from their own actions, actions of others, and developments in broader social (and physical) environments. The associated processes are non-ergodic in the sense that statistical patterns derived from past developments do not provide reliable guidelines (or predictions) concerning future developments.[2]

[1] Here substantive rationality is a useful simplifying assumption (clearly inconsistent with uncertainty as defined here) that can illustrate conformity effects. Classical game theory can also illustrate the definition of an institution (informal or formal) by representing "the rules of the game" in myriad social contexts.
[2] Knight (1921) distinguishes risk from uncertainty. On non-ergodic processes, see North (2005).

The Oxford dictionary defines a *narrative* as "A spoken or written account of connected events; a story." Narratives imply conceptions of causal relationships. In this sense, narratives offer (usually quite informal and sometimes ambiguous) models of cause and effect.

Punctuated equilibria are outcomes of dynamic processes that generate long periods of relative stability that periodically succumb to rapid bursts of change, which usually transition to a new but substantially different (punctuated) equilibrium. The initial concept applied to biological evolution. Gould and Eldridge (1977) assert that evolution exhibits relatively long periods of stability interrupted by rapid bursts of change (e.g., the mass extinction of the dinosaurs). More recently Baumgartner and Jones (1993) applied this concept to the dynamics of policymaking.[3] Punctuated equilibria can be either self-reinforcing or self-undermining, reflecting the evolution of their capacity to withstand external and internal shocks.

This paper assumes *boundedly rational* individuals: individuals possess limited and costly cognition, yet they pursue goal-oriented behavior in the sense that responses appear better (by an agent's own reckoning) than perceived feasible alternatives. Goals may be either material (money, goods and services, time, health) or social (status, concern for others often manifested as reciprocal inclinations; concern for procedure).[4] Agents adjust their understandings changing environments with trial-and error-learning.

Two additional concepts relate uncertainty to boundedly, rational agents. First, the *competence-difficulty gap* (CD gap) represents the gap "between an agent's competence and the difficulty of the decision problem to be solved (Heiner, 1983, 562)." Along similar lines, Brian Arthur's (1992, 1994) *problem-complexity boundary* represents a conceptual delineation within

[3] See also True, Baumgartner, and Jones (2007).

[4] Bounded rationality facilitates a minimalist conception of rationality, as in Ostrom (1996) and Gintis (2009). On social preference theory, see Fehr and Fischbacher (2002), Bowles (2004), and Bowles and Polanía-Reyes (2012).

which problems are sufficiently simple for standard (substantive) rational deduction to operate successfully; outside this boundary bounded rationality applies. Relating the two concepts, outside of the complexity boundary, the CD-gap precludes substantive rational analysis. This distinction leads to the Kahneman-Tversky (1979) notion of cognitive processes.

Basic cognitive processes: Daniel Kahneman (2003) asserts that people use two mental systems to process information and create judgments: a reactive, habitual, impressionable, doubt-suppressing, and essentially effortless intuitive system (S1); and a slow, deliberate, reflective, and effortful reasoning system (S2). S1 processes generate impressions by utilizing prior knowledge, responding to language (with association, or quick approval or disapproval), and incorporating learning with repetition over time. With practice, physical motions (learning to ride a bicycle) or mental processes (simple addition) can become intuitive. By contrast, S2 processes involve reasoned deductions, such as solutions to differential equations. Furthermore, S2 can sometimes monitor S1 processes and make corrections. The accessibility, or "ease with which mental contents come to mind" (Kahneman 2003, 1452), distinguishes S1 from S2 processes.[5] Learning can increase the accessibility of initially difficult processes (e.g., bike riding, addition) moving them toward the S1 end of an accessibility spectrum and, in the process, improve the accuracy of S1 procedures (a key role of education). Changes in the non-ergodic environment can, however, render prior lessons inappropriate to current circumstances.

Two additional properties of S1 processes suggest the importance of narratives and social context. First, *reference dependence* means that intuitive evaluations respond to changes from

[5] Accessibility is a spectrum with S1 "natural assessments" such as physical properties of size or loudness as well as abstractions like similarity or mood (1453) at one end, with S2 conscious deliberation at the other. Selten (1978, 1990) posits three categories of thought: routine, imagination, and reasoning. Imagination—that is, forward-looking visualization of prospects or possibilities—could occupy the middle of the S1–S2 spectrum. For an insightful discussion of imagination in economics, see Bronk (2009).

some reference point, such as the status quo.[6] For example, people show markedly different preferences toward risk depending on whether the same probabilistic outcome is described in terms of losses from a high reference point or gains from a low one (Kahneman and Tversky 1979; Tversky and Kahneman 1992).[7] *Framing effects* are conceptual responses to the manner in which ideas or events are presented. Pollsters, for example, find that the wording of questions about specific policies influences rates of policy approval (Schuman and Presser 1981). Narratives frame causal relationships with reference to specific starting points and/or conditions.

Heuristics and Mental Models

Individuals combine S1 and S2 processes, in varying degrees depending on prior experience and the difficulty of the problem at hand, to form judgments—outcomes of cognitive processes. Judgments involve heuristics and the broader concept of mental models.

Heuristics are easily accessible, largely S1 processes that combine inputs from prior experience and learning—typically in a reference-dependent or framed fashion—to produce impressionistic judgments. For example, an *attribute substitution* heuristic involves judging people, objects, or events on the basis of their most accessible (salient) attributes, such as a person's looks.[8] Attribute substitution lowers the threshold number of observations needed to form a judgment, economizing on cognitive effort. For example, knowing that John is a lawyer, makes it easy to associate his actions with aggressiveness. Yet if one assumed erroneously that he is a first-grade teacher, this conclusion might require more observations or never even arise. A

---

[6] Traditional theory presupposes reference-independent preferences that only evaluate (or rank) outcomes. Tracing origins back to Bernoulli (1738), Kahneman labels this utility concept "Bernoulli's error" (2003, 1455).

[7] "Indeed, the incorrect assumption that initial endowments do not matter is the basis of Coase's theorem and of its multiple applications" (Kahneman 2003, 1457, citing Kahneman, Knetsch, and Thaler 1990). Thaler's (1980) experiments indicate an endowment effect whereby people would sell a mug they own for twice the price they would pay to get it.

[8] Agents who use an "affect" heuristic base their judgments on initial like or dislike (Kahneman 2003).

second type of heuristic, an *affect heuristic* uses initial impressions of liking or disliking to influence (or generate) subsequent evaluations. Although heuristics economize on cognitive costs, they can introduce systematic bias and so yield inaccurate judgments.

Mental models (MM) are conceptual representations that combine basic categories (social and physical) with cause-and-effect relationships. MM combine S1 and S2 processes, including emergent heuristics such as attribute substitution, to form judgments. Over time, inputs from prior experience, learning, intuitive, inductive, and deductive reasoning combine to form (or congeal into) MM.

An Illustrative Model of Evolutionary Bounded Rationality

Herbert Simon's (1955) model of boundedly-rational behavior uses evolutionary logic to illustrate the use of mental models by boundedly rational agents. There are three basic elements: a set of alternatives (A) that specifies all possible actions or strategies of all members of the relevant interactive group within a specific social context; a set (S) of possible future states of the world (outcomes) that follow from combinations of actions; and a payoff function V(*s*), where $s \in S$ specifies a particular outcome.[9]

Boundedly rational agents, operating within the confines of a CD gap however, only perceive (consider) subset $A^o \in A$. They do not know all possible cause-effect scenarios $A \rightarrow S$ or even the related probabilities. Even so, they learn adaptively by, within limits, by observing certain outcomes from their own or others' behavior. On this basis, agents begin with prior conceptions (inherited from past experience) that allow a coarse mapping of relations between A^o and *S*. They focus effort on a subset of considered actions $A_g^o \in A$ that they expect to generate

[9] In game-theoretic terms, A and S respectively depict all strategic choices that emerge from a single node or information set in an extensive-form classical game and all outcomes that could follow.

good future states $S_g^s \in S$, defined as states that generate $V(s) = 1$ in the simplified value set $V(s) \in \{-1, 0, 1\}$, whose numbers respectively correspond to an improvement, no change, and a decline.[10] Agents then generate time-sensitive cognitive mappings $A_{1g}^o \rightarrow S_{1g}^s$ to guide their choice of actions. Such mappings can represent mental models, and elements of the mappings, such as attribution of a specific relation (→), can represent heuristics.

Heuristics and MM incorporate, respond to and/or become narratives. Regarding heuristics, narratives employ and facilitate attribute substitution and affect heuristics. For example, stories often stress specific characteristics of individuals (making them salient) in a manner that evokes judgment (e.g., big bad wolf). Narratives influence mental models not only by shaping and more broadly by illustrating categories (e.g., varying types of protagonists), but also by portraying causal relationships (stories). Narratives can, themselves, become a type of mental model. The first assertion follows.

Assertion 1: *Heuristics and, by extension, mental models respond to shared narratives in a fashion that can generate conformity of belief and action.*

Shared narratives transmit heuristics and mental models (or basic components) across groups; they convey conceptions of social (and other) categories along with cause-effect relationships across various groupings of individuals.[11] Narratives can lower cognition costs; narratives offer simple forms of attribute substitution and their stories of actors, cause, and effect provide simple and sometimes attractive representations of causality, often leading to cognitive absorption by multiple parties. Ensuing conformity effects arise from and create cognitive positive-feedback loops in which shared impressions reinforce each other. Shared narratives thus infuse MM and

[10] This is Simon's "satisficing" function (the goal is to do satisfactorily). The model could incorporate a more general value function.
[11] Game theory facilitates creating and transmitting formalized mental models that rely heavily (though not exclusively) on S2 deductive processes.

with sufficient positive feedback, specific understandings, narratives, judgments may permeate a group.

A relatively simple classical multiplayer game of assurance can illustrate. Figure 1 (see Appendix A) offers an example focused on belief in the possibility of substantive reform—a type of collective-action problem.[12] The game illustrates how interactions between expectations and motivation generate a critical-mass support threshold that distinguishes propensities towards wide-spread support from wide-spread quiescence (a type of free-riding). The game takes the point of view of an individual who faces a strategic choice between supporting reform (R) and staying quiet (Q), where her utility payoff (social and material) depends on her beliefs concerning how many others will choose R. The model has three Nash equilibria: one at full support for reform; one at full quiescence; and an unstable internal equilibrium at the critical-mass tipping point level of support (n^*). If initial expectations of the level of support for reform (n) suggest an $n < n^*$, then momentum pushes towards the full quiescence equilibrium, and vice versa. Narratives of support and opposition can play a critical role in shaping such expectations.

The next assertion applies the logic of a critical-mass tipping point to adaptive learning by boundedly rational agents.

Assertion 2: *Mental models follow the dynamics of punctuated equilibrium processes.*

A punctuated-equilibrium dynamic arises from the learning properties of mental models. MM facilitate two basic types of learning. *Internal hypothesis testing* involves thought experiments that utilize prior expectations formed by existing mental models. Informational feedback confirms or refutes such hypotheses. For example, an entrepreneur employs a business plan (a type of MM) to guide an initial investment strategy, perhaps expecting a first-year profit of 5%.

[12] An evolutionary game that interprets strategies C and D as programmed phenotypes, generates a similar result.

At the year's end, revenue data either confirms or rejects the hypothesis. Internal hypothesis testing tends to improve in stable environments and utilizes relatively low cognition costs, compared to the alternative.

The considerably less frequent *reevaluative learning* substantially adjusts or replaces MM in response to significant adverse feedback. Because they economize on cognitive effort, reevaluating MM incurs high cognition costs. Agents thus have incentives to retain already formulated models; they become "wedded to" specific conceptual frameworks. Yet the circumstances that prompt internal hypothesis testing change more often than underlying mental models. The durability of mental models and their reliance on accumulated prior assessments thus jointly generate properties of path dependence: change must come from existing states of affairs (concepts), and prior conditions influence subsequent outcomes (judgements; MM). More precisely, the evolutionary dynamics of reevaluative learning create punctuated equilibria. During the stable phase, MM serve as foundations for hypothesis testing, with incremental adjustment in response to associated findings. A sufficient volume of inconsistent findings (e.g. a dramatic surprise), however, undermines MM, punctuating prior equilibria (Denzau and North 1994).[13] After the fifth successive year of losses, the entrepreneur needs to revise her entire conception of business prospects: reevaluative learning.[14]

In terms of game-theoretic modeling, the critical-mass concept can illustrate the path-dependency property of MM as cognitive punctuated equilibrium. In Figure 1, if initial expectations of support for reform (*n*) fall below the critical-mass tipping point *n**, individuals find quiescence more attractive that reform activity. Observations will tend to reinforce

13 Unlike pure Bayesian expectations, sufficient observation of problems or inconsistencies can lead to complete reevaluation (Kahneman 2003). Mental models are analogous to Kuhn's (1970) scientific paradigms.
14 Kahneman notes that correction processes are often incomplete. "There is a continuum of theories that agents can hold and act upon without ever encountering events which lead them to change their theories" (Hahn 1987, 324; cited in North 2005, 62).

expectations; positive feedback effects will lead to an equilibrium of quiescence, both in terms of belief and action. Should, however, salient events reorient prevailing expectations, anticipated support may cross the n^* threshold—a punctuation.

In an analogous fashion, Ronald Heiner's (1983) model of a reliability condition can represent this second assertion (and the first). At face value, Heiner's model applies evolutionary logic to represent how agents who face uncertainty select among known strategies for inclusion into heuristics and by extension mental models. The key idea is that when boundedly rational agents face uncertainty, heuristics can improve average performance by excluding consideration of certain feasible actions when the likelihood of identifying appropriate circumstances for their use is too low. For example, a potential shoplifter may avoid considering such activity because it is too difficult to identify circumstances in which one would not get caught. Heuristics thus include only strategies that meet a basic reliability threshold in the presence of uncertainty. Absent underlying information on relevant probabilities, reliability follows an evolutionary selection logic: strategies that, on the basis of prior experience, appear to do well remain in choice sets (see Appendix B for details). The model may be extended to represent the inclusion of heuristics and narratives into mental models. In either case, once established, the threshold implies that heuristics and MM will persist until events generate two conditions: i) sufficient problems (or inconsistencies) with the initial MM; and ii) a conception of one or more alternative strategies (heuristics, MM) that can cross a reliability threshold. Meeting both conditions may be quite infrequent.

Assertion 1's conformity effects of shared narratives can now be extended to generate the next assertion.

Assertion 3: *Institutions are a type of shared mental model that convey basic understandings across uncertain environments.*

Communication fosters sharing of the categories and causal relationships that establish mental models, often lowering costs of cognition. *Shared mental models* incorporate common vocabulary, categories, perceived patterns, and causal relationships. According to Denzau and North (1994), there are two basic types of shared mental models: ideologies and institutions. Narratives play a central role in both types.

Ideologies are shared MM that address uncertainty by relying on narratives about causal relationships (with categories) that interpret social environments and apply normative judgments to various activities, outcomes, and ultimately to how social environment should be structured or changed. For example, a libertarian ideology might interpret government activity as largely interfering with tendencies of people to produce and exchange and thus advocate for low taxes and deregulation. The shared narratives of ideologies can generate conformity of belief (cognitive monocultures) among adherents.

Institutions, simply put, are "the rules of the game in a society," (North, 1990). More precisely, an *institution* is a combination of mutually understood and self-enforcing beliefs, decision rules, conventions, social norms, or formal rules that jointly specify or prescribe behavioral regularities in specific or varied social contexts (Ferguson 2013, 152).[15] Broadly speaking, there are two basic types of institutions. *Formal institutions* (formal rules) arise from specified collective decision-making processes and are typically, though not always, written. Examples include constitutions, legislation, and formally decided government or corporate rules,

[15] By contrast, o*rganizations* are somewhat fluid groupings of boundedly rational individuals with differing interests who pursue sets of negotiated goals and use evolving decision rules to understand and coordinate key operations (Cyert and March 1963). As relatively coherent groups, organizations operate as agents within institutional contexts. In game-theoretic terms, institutions are rules of the game that establish the context within which organizations (and individuals) act. The same entity may be regarded as both an institution and an organization (a reason why popular discourse so often mixes the two). For example, as an institution, the General Motors Corporation is a set of rules governing procedures like personnel and sales policies. As an organization, GM produces automobiles and lobbies Congress.

such as personnel policies. *Informal institutions* are conventions, norms, or other informal behavioral patterns that emerge more or less spontaneously from repeated social interactions or from informal decision procedures. *Social norms* are expected behavioral prescriptions with ethical content and unspecified but generally understood social enforcement. For example, one should not cut in line.

Institutions accomplish three related tasks that facilitate coordinating behavior and belief. First, motivation: In a most straightforward sense, institutions motivate various types of behavior by establishing (understood) incentives. For example, a tax on cigarette consumption may reduce smoking among youth (and others) who are not yet addicted. The ethical content of social norms exerts a possibly more profound and more subtle institutional influence on motivation. People who adhere to specific social norms (e.g., don't cut in line), hesitate to violate them—out of hesitation to face social sanction and/or an internal sense of discomfort; the latter effect shapes preference orderings over time. Second, institutions influence flows of information, both by generating certain types of information (e.g., the tax on cigarettes is $.50 per pack) and by specifying rules that structure information flows (e.g., the chief financial officer receives all revenue reports).

Third, as shared mental models, institutions shape cognition. They incorporate and transmit common conceptions of fundamental social categories as well as common conceptions of cause-effect relationships. Once a practice has become institutionalized, agents share an expectation (a conception) of general conformity or adherence to certain types of prescribed behavior that apply to specific categories of individuals—all within specific social contexts or circumstances. For example, a social norm can prescribe that when an elderly person enters a full bus, a young person should offer a seat. This prescription influences cognition by specifying categories (young, elderly), applicable conditions (the seats are full), and expected action—a causal relationship. Riders on the bus may then predict activity that would defy prediction in the absence of the norm. The riders possess a shared understanding of a specific causal relationship applicable to specific categories: a shared mental model. Similarly formal rules, when institutionalized (i.e., when commonly understood and expected), exert a similar conceptual and motivational influence. Shoppers understand that if a manager observes them pocketing an item from a store, they are likely to be reported and prosecuted.

The fourth assertion follows.

Assertion 4: *By enabling boundedly rational actors to manage uncertainty, institutions effectively choreograph social activity.*

Because fundamental uncertainty renders standard risk calculus unreliable; because boundedly rational agents face limited cognition, high costs to deductive reasoning, CD gaps, and associated problem-complexity boundaries, agents need cognitive mechanisms that involve considerably lower costs and broader applicability than individual S2 deductive reasoning. Shared mental models, notably social norms and formal institutions, play this role.

More precisely, the cognitive properties of institutions generate a form of *social choreography* (Gintis, 2009). In complex social environments, commonly understood institutional prescriptions shape motivation, convey information, and influence (frame) beliefs related to specific causal relationships in specific social contexts.[16] Institutional signals thus generate conjectures about behavior in relevant contexts. The ethical dimensions of social norms reinforce these signals, influencing preference orderings by creating internalized values (ethical deontics) that promote adherence to associated prescriptions. Shared understandings, then promote behavioral adherence: a self-reinforcing positive feedback loop between understanding

[16] See Crawford and Ostrom (2005) on the syntax of institutional prescriptions.

and behavior.[17] Multiple agents may then approach specific social contexts with common priors that not only render the behavior of other agents at least somewhat predictable, but that also coordinate belief and action, generating a type of conformity: social choreography.

In the previously noted bus example, a social norm coordinates belief and activity among riders. Similarly, a formal prohibition of shoplifting, when institutionalized (likely with supporting social norms), has a comparable impact: Absent such a shared understanding, as might for example be the case in the midst of a civil war, neither shoppers nor owners could reliably predict shopping behavior. Institutions thus reduce uncertainty by coordinating understandings and activity; they reduce the CD gaps of numerous agents by effectively ruling out myriad forms of defection (anti-social behavior), providing a source of predictability. Institutions thereby resolve multiple collective-action problems, especially ones related to social coordination and enforcement (though likely creating new problems in the process).

An Epistemic Game Model of Shared Understandings

Epistemic game theory can model shared and limited understandings among boundedly rational agents and, by extension, how institutions choreograph their social activity. In an epistemic game there is a large set of possible states of the world (S). Each agent has a knowledge partition (P_i) that represents that agent's ability to distinguish among possible states of the world. For example, before I look out the window in the morning, my knowledge partition may not distinguish between a rainy day and a sunny day. Accordingly, agents may not be able to distinguish the actual current state of the world (ω_c) from other possible states. Each agent also possesses a possibility operator ($P_i\omega_c$) that specifies the states that she considers possible (maybe

[17] Bronk (2013, 345) refers to essentially the same phenomenon as *performativity*.

rain; maybe sun). For any state ω, each agent forms a subjective prior ($\rho_i(\omega)$) that specifies her conjectures about the strategies of others in state ω (some people may carry umbrellas to work).

An epistemic game proceeds as follows. An event (or nature) sends a signal (γ) to players who, on that basis, share a conjecture with greater or lesser degrees of precision. In a fully correlated case, the signal specifies the same strategy for all players. In a minimal case, players know only that everyone received some kind of signal. Intermediate cases are possible.[18] Signals influence subsequent actions by first affecting agents' subjective priors. The type and precision of signal specifies the degree to which agents share common understandings (see Appendix C for a simple epistemic game).

Gintis (2009) argues that epistemic games foster a concept of *social epistemology*, whereby institutions (he focuses on norms and conventions) influence cognition, information, and motivation. Institutional prescriptions provide the epistemic signals (γ) that correlate understandings and thereby coordinate action, illustrating social choreography.

Social Choreography, Conformity, and Predictability

Previous arguments have established that individuals form and use heuristics and associated mental models to navigate uncertainty. MM reflect (emerge from) prior experience and utilize, incorporate, and even become narratives. Because existing MM allow agents to test hypotheses at a low cognitive cost, whereas re-evaluative learning is costly, mental models follow dynamics of punctuated equilibria. Communication, moreover, facilitates sharing of MM, usually lowering cognitive costs. Institutions, as a major type of shared MM, shape motivation, information, and cognition among multiple agents. In so doing, institutions choreograph belief and activity across

[18] Gintis' (2009) discussion draws on Auman's (1987) concept of correlated equilibria.

groups or populations. Institutional social choreography, with its conceptual foundation in shared mental models, also follows punctuated-equilibrium dynamics.

During the stable phase of a dynamic cycle, the social choreography of mutually understood institutional prescriptions renders many social activities in multiple social contexts relatively predictable. The consequent decline in the CD gap enables boundedly rational agents to navigate such terrain utilizing some S2 reasoning, but such reasoning relies on S1 heuristics and the framing effects of social categories and causal relationships that emerge from social norms and formal institutions. Associated narratives and ideologies reinforce belief and behavior, with self-reinforcing feedback, often enhancing tendencies toward conformity of belief and practice—social coordination.[19] In the prior bus example, the social norm of offering a seat may be reinforced by narratives that portray an act of placing one's own comfort above that of others as selfish. Joint adherence to the norm generates conformity (a type of monoculture), arguably a benign one in this case. More ominously, narratives of racial difference have supported monocultures that generate and practice norms of exclusion.[20] Referring to the economics of finance, institutional practices that establish the authority of bond-rating agencies have generated a conformity of perspective that, prior to the 2008 crisis, failed to seriously consider correlations among various types of financial risk that, at the time, appeared to offset each other (Bronk 2013, 2014).

Furthermore, during the stable phase of institutional evolution, significant change—that is significant reform, abolition or replacement of institutions—requires altering shared mental models; it requires simultaneous reevaluative learning that is at least transmitted (if not initiated)

[19] Institutional equilibria can be either self-reinforcing or self-undermining, reflecting the evolution of their capacity to withstand external and internal shocks (Greif 2006). One may model self-reinforcing and self-undermining dynamics as respectively increasing or diminishing basins of attraction. The size of such a basin at a point of time represents the stability of its associated punctuated equilibrium.

[20] An example of institutions creating substantial collective-action problems.

across broad groups. Generally speaking, shared re-evaluation requires most of the following characteristics: significant occurrence of salient inconsistencies; agents who experience such inconsistencies who occupy positions (social, political, and/or economic) that facilitate transmission; communication of such inconsistencies or problems to a sufficiently large audience of relevant players; an imitative process of collective reassessment.[21] One may wonder, for example, whether the 2008 financial crisis generated sufficient shared inconsistency with previously reigning understandings of risk to promote significant reevaluation among financial analysts. Narratives about the nature of the crisis—its causes, its impacts—have and will continue to shape prospects for such re-evaluation.[22]

By contrast, during the punctuation phase of an institutional dynamic, the breakdown of prior social choreography unravels the predictability of relevant social encounters in relevant social contexts. The problem complexity boundary shrinks dramatically as the CD gap grows. During a political revolution, for example, the predictability of prior processes (e.g., operation of the legal system) collapses. During such periods agents face fundamental uncertainty in many social encounters, and relatively small events may generate cascades of shifting belief and activity.

[21] Social network theory can model both the influence of institutions or dominant narratives and the process of punctuation of associated equilibrium, via the concept of information cascades. For a thorough discussion of social network theory, see Jackson (2008). For a discussion of information cascades, see Watts (2002). For basic applications of social networks and information cascades to collective-action problems, see Ferguson (2013).

[22] Significantly reforming institutions and associated practices involves resolving large collective-action problems (see Ferguson, 2013, Chapter 12; Ferguson 2017).

Conclusion: A Few Thoughts on Institutional Change

As shared mental models, norms and formal institutions convey basic understandings and conceptual frameworks, often utilizing shared narratives, across groups of individuals. These understandings and frameworks help boundedly rational agents, agents who face significant CD gaps outside of relevant problem-complexity boundaries, navigate uncertainty—despite the costs and limitations of S2 deductive mental processing. Institutional prescriptions specify expected behaviors that apply to specific categories of individuals operating within specific social contexts. In so doing, institutions effectively choreograph both belief and activity. The associated processes, moreover, follow a punctuated-equilibrium dynamic that emerges from positive-feedback loops of self-reinforcing belief and behavior, along with associated path dependency. During relatively long periods of institutional stability agents can often navigate myriad social encounters because institutional prescriptions render behavior in complex environments relatively predictable; social choreography facilitates predictability.

On the other hand, sufficiently salient inconsistencies (or problems) with existing institutional shared mental models, if held among a sufficient number of well-placed agents with sufficient transmission across relevant populations, can punctuate institutional equilibria. Significant institutional change follows. Punctuation, responding to cascades of shifting understandings, significantly alters, abolishes, and or replaces extant institutions. During such periods, CD gaps expand markedly and prior predictability breaks down.

Two political economy implications are worth noting. First, leadership and power affect the content, transmission and potential adjustment of shared understandings. One manifestation of leadership and its associated power is the ability to shape shared narratives and thereby influence or manipulate shared understandings, shared heuristics, shared narratives, and ultimately shared mental models. Moreover, leadership itself is path-dependent. Second, significant reform of institutions requires dislodging their punctuated equilibria. Institutional reform thus requires resolving substantial collective-action problems of displacing shared mental models and their corresponding social choreography of belief and behavior. Reformers must not only overcome opposition from (often powerful) parties who benefit from the status quo and basic problems of coordinating oppositional activity, they also face daunting collective-action problems related to achieving a critical mass of shifting conceptions among boundedly rational agents who face CD gaps in uncertain environments.

References

Arthur, W. Brian (1992) "On Learning and Adaptation in the Economy," Santa Fe Institute Working Paper 92-070038.

______ (1994) "Inductive Reasoning and Bounded Rationality," *American Economic Review* 84(2), 406–11.

Aumann, Robert J. (1987), "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica* 55(1), 1–18.

Baumgartner, Frank R., and Bryan D. Jones (1993), *Agendas and Instability in American Politics*, Chicago: University of Chicago Press.

Bowles, Samuel (2004), *Microeconomics: Behavior, Institutions, and Evolution*, Princeton, NJ: Princeton University Press.

Bowles, Samuel, and Sandra Polanía-Reyes (2012), "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature* 50(2), 1–57.

Bronk, Richard (2009), *The Romantic Economist: Imagination in Economics*, Cambridge: Cambridge University Press.

______ (2013) "Hayek on the wisdom of prices: A reassessment," *Erasmus Journal for Philosophy and Economics, Volume 6, Issue 1, 82-107.*

______ (2014) "Reflexivity unpacked: performativity, uncertainty and analytical monocultures," *Journal of Economic Methodology*, 20:4, 343-349.

Conlisk, John (1996), "Why Bounded Rationality?" *Journal of Economic Literature*, XXXIV, 669-700.

Crawford, Sue and Elinor Ostrom (2005), "A Grammar of Institutions," in in Elinor Ostrom, ed. *Understanding Institutional Diversity*, Princeton, NJ: Princeton University Press, 137-74.

Cyert, Richard M., and James G. March (1963), *A Behavioral Theory of the Firm*, Englewood Cliffs, NJ: Prentice Hall.

Denzau, Arthur T. and Douglass C. North (1994), "Shared Mental Models: Ideologies and Institutions," *Kyklos* (47), 3-31.

Dixit, Avinash, Susan Skeath, and David H. Reiley Jr. (2009), *Games of Strategy*, Third Edition, New York: Norton.

Fehr and Fischbacher (2002), "Why Social Preferences Matter—The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives," *Economic Journal* 112(478), C1–C33.

Ferguson, William D. (2013), *Collective Action and Exchange: A Game-Theoretic Approach to Contemporary Political Economy*, Stanford, CA: Stanford University Press.

______ (2017) "The Political Economy of Collective Action, Radical Reform, and Revolution," SSRN Working Paper 3405446.

Gintis, Herbert (2009a), *Game Theory Evolving: A Problem-Centered Introduction to Strategic Interaction*, 2nd ed. Princeton, N.J.: Princeton University Press.

_____ (2009b), *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*, Princeton, N.J.: Princeton University Press.

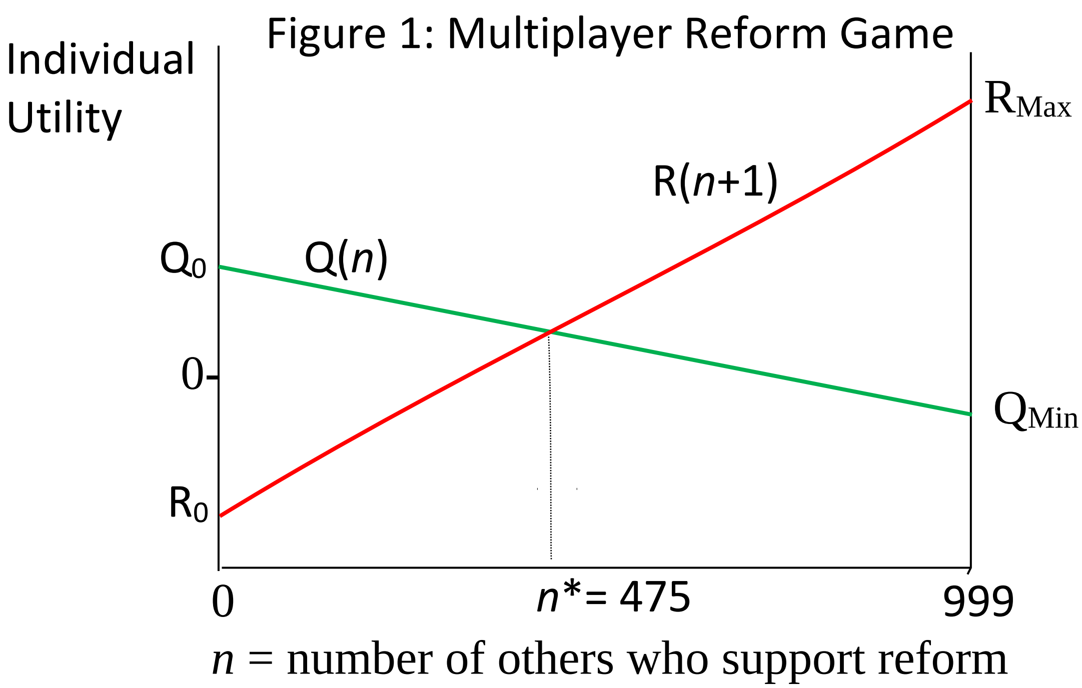
Gould, Steven J and Niles Eldredge (1977) "Punctuated equilibria: the tempo and mode of evolution reconsidered," *Paleobiology* 3(2), 115-151.

Greif, Avner (2006a), *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*, Cambridge: Cambridge University Press.

Hahn, Frank (1987), "Information, Dynamics, and Equilibrium," *Scottish Journal of Political Economy* 102(5), 912–50.
Heiner, Ronald A. (1983), "The Origin of Predictable Behavior," *American Economic Review*, 73(4), 560-595.
Jackson, Matthew O. (2008), *Social and Economic Networks*, Princeton, NJ: Princeton University Press.
Kahneman, Daniel (2003), "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review* 93(5), 1449-1475.
Kahneman, Daniel, and Amos Tversky (1979), "Prospect Theory: An Analysis of Decisions under Risk," *Econometrica* 47(2), 263–91.
Kahneman, Daniel, Jack Knetsch, and Richard Thaler (1990) "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy* 96(8), 1325-48.
Knight, Frank (1921) *Risk, Uncertainty and Profit*, Boston, MA: Houghton Mifflin.
Kuhn, Thomas (1970), *The Structure of Scientific Revolution*, 2nd ed., Chicago: University of Chicago Press.
North, Douglass C. (1990), *Institutions, Institutional Change and Economic Performance,* Cambridge: Cambridge University Press.
______ (2005), *Understanding the Process of Economic Change*, Princeton, NJ: Princeton University Press.
Ostrom, Elinor (1996), "A Behavioral Approach to the Rational Choice Theory of Collective Action," *American Political Science Review* 92(1), 1–22.
Ostrom, Elinor, and Sue Crawford (2005), "Classifying Rules," in Elinor Ostrom, *Understanding Institutional Diversity*, Princeton, NJ: Princeton University Press, 186–215.
Schuman, Howard and Stanley Presser (1981) *Questions & Answers in Attitude Surveys*, San Diego, CA: Academic Press.
Selten, Reinhard (1978), "The Chain Store Paradox," *Theory and Decision* 9(2), 127-59.
______ (1990), "Bounded Rationality," *Journal of Institutional and Theoretical Economics,* 146(4), 649-658.
Simon, Herbert (1955) "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69(1), 99–118.
Thaler, Richard H. (1980), "Towards a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization* 1(1), 36–90.
True, James L., Bryan D. Jones, and Frank. R. Baumgartner (2007), "Punctuated-Equilibrium Theory: Explaining Stability and Change in Public Policymaking," in Paul A. Sabatier, ed. *Theories of the Policy Process,* Second Edition, Boulder, CO: Westview Press, 155-187.
Tversky, Amos, and Daniel Kahneman (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5(4), 297–323.
Watts, Duncan J (2002) A simple model of global cascades on random networks, *Proceedings of the Natural Academy of Sciences* 99(9), 5766–71.

Appendices

A. Multi-Player Assurance: Expectations and Conformity



There are 1000 members of the relevant group. The model depicts an individual's choice between R (support reform) and Q (remain quiescent). On the horizontal axis, the variable *n*, represents the number of <u>other</u> people (out of 999) who the individual in question expects to choose R. The vertical axis represents this individual's net utility payoff—a combination of material and social gains or losses—that arise from either activity. A net payoff may be either positive or negative (hence 0 is not at the bottom of the vertical axis).

Because reform activity involves solidarity and conformity effects, the R(*n*) function slopes upward; higher expected support among others increases an individual's payoff to R. In equation form:

(1) $R(n + 1) = R_0 + r(n + 1)$

where R_0, Figure 1's left-hand intercept, specifies the individual's expected utility from choosing R if no others do so (at $n = 0$); and r (the slope coefficient) represents the expected marginal utility gain ($\partial R/\partial n$) from each expected contribution to reform by somebody else.[23] Note that $R_0 < 0$: the expected payoff to activity R is negative if nobody else does so (i.e., bad consequences, such as social ostracism or jail time). At high levels of expected participation from others, the individual's net payoff to R becomes noticeably positive, as shown by right-hand intercept R_{Max}. A similar (mirror image) logic applies to staying quiet as illustrated by the Q(n) function.

(2) $Q(n) = Q_0 - qn$

At intercept Q_0, where $n = 0$, the payoff to staying quiet is high. But this expected payoff declines steadily as expected support for reform increases (at the rate $-q$), becoming negative at right-hand intercept Q_{Min}.

The model has three Nash equilibria, two stable equilibria at the extremes of no support ($n = 0$ and the individual chooses Q) and full support for reform ($n = 999$ and the individual chooses R), as well as an unstable internal equilibrium at point n^*, where the two lines cross, signifying equal payoffs to either activity.[24] At $n = 0$, the payoff from Q exceeds that from R. If all members of the relevant group expect such a complete lack of support, they will also choose Q. Likewise at the full-support equilibrium, the return to R exceeds that from Q. Consequently, a shared expectation of full support offers group members an incentive to contribute to reform. In either extreme case, there is no inherent tendency for change, no individual incentive to switch strategies: a Nash equilibrium. Moreover, should a small number (perhaps mistakenly) choose

[23] The term R(n + 1) signifies that if (and only if) the individual chooses R, then there are n+1 participators.
[24] Point n^* is a Nash equilibrium because, given the expected behavior of others, the individual has no incentive to alter her behavior.

the opposite strategy, say R in any situation when $n < n^*$, payoff incentives will tend to induce them to switch back to Q; hence the outcome returns (or moves) to the ($n = 0$, Q) equilibrium.

Either stable equilibrium outcome, then, is possible and, absent some external change, is sustainable in at least the short run. The internal equilibrium at $n = n^*$ is unstable because a small deviation in either direction induces continued deviation in the same direction—reflecting a positive-feedback dynamic. As such, n^* represents a critical-mass threshold (a tipping point) of expected support that often accompanies reform processes. A general expectation of wide-spread support tends to motivate individuals to actually support reform in a self-reinforcing fashion, and vice versa.[25] The model suggests path-dependence, often accompanied by the related phenomenon of self-fulfilling expectations. Initial conditions, specifically concerning which side of n^* a group starts on, influence subsequent equilibrium outcomes.

B. The Evolution of Reliable Strategies, Heuristics, and Mental Models26

Here is a model (based largely on Heiner (1983)) that uses evolutionary logic to show why heuristics improve outcomes for boundedly rational agents facing uncertainty. We use Simon's concepts of *A*, *S*, and *V* (see p.6)—noting that set *S* (possible future states) includes current states of the world (outcomes of prior actions) as conditions under which strategies are chosen (i.e., $A(S_0) \rightarrow S_1$). The model's core logic: heuristics improve average performance by eliminating an agent's consideration of certain feasible actions when the likelihood of identifying appropriate circumstances for use of such actions is too low. The model's applications include acquisition of

[25] One could adjust the model to allow for some agents who will always, independent of the activity of others, either support or oppose reform. In such cases, the model's stable equilibria occur where activities R and Q achieve their minimum and maximum levels.
[26] Appendix B is taken, practically verbatim, from Ferguson 2013, pp. 132-135.

information, selection of heuristics within mental models, and reevaluative learning that adjusts or replaces individual or shared mental models, including social norms and formal institutions.

Boundedly rational agents face a CD gap, the extent of which depends on relationships between environmental complexity and agents' cognitive abilities. The vast set S (all possible states that exist within or could arise from the current environment) represents such complexity. Cognitive abilities (Θ) indicate an agent's "competence in deciphering relationships between its behavior and the environment" (Heiner 1983, 564). Hence

(3) CD gap = $\upsilon(S, \Theta)$,

where υ indicates the extent of uncertainty; ($\partial\upsilon/\partial S > 0$) and ($\partial\upsilon/\partial\Theta < 0$). Increasing either the number of agents or the set of feasible alternatives (A) increases the size of S exponentially.[27] Population growth and technological change exponentially increase environmental complexity and hence the CD gap because these trends increase both the number of interacting agents and the size of set A.[28] Even though some technologies, such as binoculars or computers, counteract this effect by increasing perception and/or cognition, economic development necessarily engenders rising complexity (North 1990).

To illustrate the model, consider a bargaining encounter between a boundedly rational worker and a firm. The firm offers the worker wage w. The worker has a prior set of reasonably understood actions $A_1^o \in A$. Suppose A_1^o includes two strategies, a and a', respectively bargain hard (ask for 5% above the offer) and bargain easy (accept the offer). Consider a new (mutant) action a'': bargain really hard (ask for 10% above the offer). There exists a set of good states (circumstances) $S_g(a'') \in S$ under which asking for 10% would improve outcomes: $V(a''; S_g) = 1$.

[27] In a one-time, two-player, two-strategy simultaneous game, there are 2^2 elements in *S*—the four cells of the game matrix. In an otherwise similar three-player, three-strategy game, there are 3^3 = 27 cells; 4^4 = 256 and so on.
[28] For an insightful argument on how technical change increases uncertainty, see Bronk (2009).

For example, a type-E (easy) firm might not want to risk losing the worker, but other firms would rather hire a replacement. Define probability ρ_g as the probability that a good situation (state) for using a'' actually exists at an appropriate time. Defining $S_g(a'')$ as the set of all possible states that are "good" for using a'' (so that $V(a''; S_g(a'')) = 1$), $\rho_g(a'', S) = S_g(a'')/S$.

Because they face a CD gap, however, agents do not necessarily know when circumstances $S_g(a'')$ will occur or even the value of ρ_g. Still, there exists in the world a probability—which is conditional on the level of uncertainty ($\rho_r(a''; \upsilon)$)—that the agent will, even with limited information and cognition, choose a'' under the "right" circumstances. In the bargaining example, ρ_r specifies the probability that the agent correctly identifies a situation (employer is type E) when asking for 10% (a'') would actually improve her outcomes. Note that uncertainty lowers ρ_r ($\partial\rho_r/\partial\upsilon < 0$). Identifying the employer's type may be difficult, especially if concealing type improves the employer's prospects. Now define $\rho_w(a''; \upsilon)$ as the conditional probability that the agent picks the "wrong" situation for using a''(employer is not type E).[29] Again, the boundedly rational agent does not know either $\rho_r(a''; \upsilon)$ or $\rho_w(a''; \upsilon)$. Note further that $\rho_r(\upsilon)$ and $\rho_w(\upsilon)$ need not sum to 1.[30]

On this basis, one can show conditions under which considering strategy a'' (i.e., including a'' in the worker's contemplated choice set) could improve the worker's outcomes. Define $g(a'', S)$ as the average gain from using a'' at the "right" time (i.e., in a good state $S_g(a'')$). Likewise, $l(a'', S)$ represents the average loss from choosing the "wrong" circumstance. Equation (4) specifies the condition under which adding action a'' to the list of considered actions (A^o) improves an agent's likely outcomes:

[29] Heiner relates $1 - \rho_r$ and ρ_w to Type I and Type II errors in statistical estimations (1983, 565, note 17).
[30] An agent that knows ρ_g, $\rho_r(\upsilon)$, and $\rho_w(\upsilon)$ faces risk, not uncertainty as in a classical substantive rationality model. See Exercise 6.4.

(4) $g(a'', S) \times \rho_r(a''; \upsilon) \times \rho_g(a'', S) > l(a'', S) \times \rho_w(a''; \upsilon) \times (1 - \rho_g(a'', S))$.

The left-hand side of (4) shows the expected gain from a "right" decision on *a″*: the gain from making a right decision, multiplied by the probability that the agent will make such a decision, multiplied by the probability that a good state will actually arise (or exist) at the right moment. Analogously, the right-hand side shows the expected loss from a "wrong" decision.

Rearranging, the *reliability condition* is:

$$(5)\quad \frac{p_r(\upsilon)}{p_w(\upsilon)} > \frac{l(a'',S)}{g(a'',S)} \times \frac{1-\rho_g(a'',S)}{\rho_g(a'',S)}.$$

The left-hand side of (5) shows the *reliability ratio*, depicting the reliability of an agent's decision about using *a″* in the presence of a CD gap; it specifies the probability that the agent makes a "right" decision to use *a″* relative to that of making a "wrong" decision. The right-hand side depicts the threshold that the ratio ρ_r/ρ_w must exceed in order for *a″* to be a *reliable* action—an action that should, on average, improve the agent's outcomes (Heiner 1983, 564–67). To simplify notation for the right-hand side, let $T(a''; S, \upsilon) = (l/g)(1 - \rho_g)/\rho_g$. Given *S* and υ, the threshold ratio $(\rho_r/\rho_w)^* = T^*$ distinguishes reliable from unreliable use of *a″*. If the existing CD gap ($\upsilon(S, \Theta)$) allows $\rho_r/\rho_w > T^*$, then *a″* is reliable: $V(a''; S) = 1$. If $\rho_r/\rho_w < T^*$, then $V(a''; S) = -1$, and *a″* is not reliable. Note that rising because uncertainty (υ) reduces ρ_r and increases ρ_w, reliability (ρ_r/ρ_w) falls.

One can use the selection criteria implied by the reliability condition, together with Simon's value function, to explain the emergence and operation of heuristics. If new action *a″* satisfies condition (5), $V(a'') = 1$; if it fails to do so, $V(a'') = -1$. From an evolutionary perspective, then, agents tend to earn higher payoffs (in terms of criteria from value function *V*) if they consider an action *a″* (i.e., add it to their set of contemplated actions A^o) only when they know enough about the environment to facilitate reliable use of that action. When an action *a″*

fails to meet the reliability condition, agents actually do better on average by using heuristics that exclude a'' from their choice set.[31] For example, there are some agents who could profit from shoplifting under the "right" circumstances, yet they do not even consider doing so because they do not trust their ability to correctly identify such circumstances.

Concerning heuristics, equation (5) implies that heuristics tend to include only actions that fit recurrent or otherwise likely circumstances.[32] Mathematically, for given values of g and l, a lower ρ_g (lower S_g/S) increases the threshold T^* that ρ_r/ρ_w must cross for satisfying (5). Equivalently, a low prevalence of good situations for using a'', makes it relatively unlikely that an agent will know, observe, or hear about such occasions; the converse applies for high prevalence of good situations. Thus evolutionary logic suggests that heuristics thus tend to include strategies that arise repeatedly in agents' prior experience, ones that they often hear about, or strategies for which good circumstances are otherwise relatively noticeable. For example, if the worker had no prior experience with type-E employers, her heuristic would likely exclude a" even though her current employer might be of type E.[33]

To complete the argument, define heuristic (h) as a set of considered actions ($a_h \in A_h^o$) that satisfies (5). Strategies that are not perceived as reliable fail to enter the heuristic. The heuristic then simplifies decisionmaking in complex environments as it significantly reduces cognition costs. Three extensions follow. First, condition (5) applies to the acquisition of information related to strategic decisions. If action a'' fails to meet the reliability condition, agents who devote resources to investigating a'' will, on average, fare worse than those who do

[31] Thus, unlike Bayesian maximization, the reliability criterion discards some possibly useful actions. "Intrinsic in behavioral rules is the ignoring of or lack of alertness to potential information" (Heiner 1983).
[32] "Uncertainty generates rules which are adapted only to likely or recurrent situations" (Heiner 1983, 567).
[33] Consumers have difficulty forming preferences over items or activities for which they lack relevant experience (Heiner 1983, 567–69). For example, insurance contracts typically address contingencies that buyers hope never to encounter.

not (unless such acquisition facilitates crossing T*). This outcome fits common sense: we do not waste energy investigating the consequences of activities that we have no intention of pursuing (even though such activities might be useful). Path dependency again arises. Second, equation (5), with slightly adjusted symbols, can illustrate the selection of heuristics for inclusion in mental models. Selection among heuristics can reflect outcomes of hypothesis testing within a given mental model. Heuristics that meet a pertinent reliability condition, by standards of the applicable MM, tend to persist. Third, reliability influences the long-term selection of mental models via processes of reevaluative learning. MM that tend to satisfy an adjusted equation (5) in relevant social contexts tend to persist. Reevaluative learning arises when altered circumstances sufficiently undermine the prior reliability of an MM. The same logic applies to shared mental models. Social norms are shared mental models which combine group heuristics that typically meet some reliability condition in their relevant social environments—given the attendant perceptions, conflicts, and distributions of power. Likewise, formal institutions emerge from analogous selection procedures that rely on collective decision making in pertinent social contexts.

It is important to realize that, despite its evolutionary logic, reliability does not imply optimality. The relevant performance criteria, for all of these applications, may reflect different types of payoffs (social or material) for different kinds of players (individuals, firms, etc.) in different contexts. Different selection pressures—such as pursuing short-term profits versus developing cooperative relationships with employees, or short- versus long-term viability—may push in distinct and conflicting directions.[34] Because multiple potentially conflicting selective pressures operate in a non-ergodic world, evolutionary selection need not generate optimality. As

[34] Selection pressure is not reducible to a single maximizable scalar.

Heiner notes, "Such a world will be a continual mixture of appropriately and inappropriately structured behavior" (1983, 569).

Overall, the reliability condition facilitates conceptualizing the selection pressures that underlie the evolution of strategies, heuristics, individual mental models, social norms, and formal institutions.

C. Example of an Epistemic Game[35]

Consider a correlated equilibrium (Aumann 1987) for a game G that has multiple equilibria. Suppose that, before playing G, each player observes some event or signal (γ) that "suggests" a specific action (pure strategy) to each player ($a_i = a_i(\gamma)$).[36] A *correlated equilibrium* exists if each player follows her own suggestion $a_i(\gamma)$, each player expects the others to follow their own suggestions ($a_j(\gamma)$), and no player can do better by unilaterally deviating from its $a_i(\gamma)$. Note that players do not necessarily know the contents of others' signals ($a_j(\gamma)$). Even so, γ induces a shared expectation among players, prior to playing G, that all will follow some suggestion from γ.[37] Event γ acts as a *correlating device*; it generates a shared expectation that directs behavioral outcomes. Note further that signal γ operates as a first move, made by nature or another player, in a larger game G+, which includes G subsequently.

For example, Figure 8.2 illustrates a game of battle in which two pure-strategy Nash equilibria exist, (U, L) and (D, R), along with a mixed strategy equilibrium in which Ann plays U and Ben plays L with probability 0.5 for expected payoff (2.5, 2.5). A simple correlated

[35] Appendix C is taken, practically verbatim, from Ferguson 2013, pp. 174-174.

[36] Observations of γ need not be independent; this is how correlated strategies differ from mixed strategies (Aumann 1987).

[37] A correlated equilibrium is a broader concept than a Nash equilibrium; the former does not require that agents know (accurately expect) others' strategies (Aumann 1987). Any rationalizable set of strategies could be a correlated equilibrium (Gintis 2009b). For an intuitive discussion of rationalizability, see Dixit, Skeath, and Reiley (2009, 157–62).

equilibrium arises as follows: with probability 0.5, event γ simultaneously signals U to Ann and L to Ben; also with probability 0.5, it signals D to Ann and R to Ben. Both players execute their indicated strategy. Expected payoffs are (3, 3). In this simple example, both players know the suggestions for the other player.[38]

Figure 8.2: Two-player game of battle

		Ben L	Ben R
Ann	U	5,1	0,0
	D	4,4	1,5

Extending this logic, the concept of an epistemic game allows us to represent informal institutions as correlating devices for complex social interactions. In an *epistemic game*, as in any game, all players have a strategy set (S_i) and there is a set Ω of all possible states of the world. Each specific state ω (in Ω) indicates a specific strategy for each player ($s_i = s_i(\omega)$). In Figure 8.2, Ann plays U and Ben plays L in one such state. For multi-player games, the set Ω is enormous. Furthermore, in epistemic games the following statements hold:

- Each agent *i* has a *knowledge partition* (P_i) that represents the degree to which she can distinguish among the different possible states of the world. Agents distinguish among cells (C_i) within partition (P_i), but within each cell, they do not know which specific state ω applies. For example, in Figure 8.2 (without γ), Ann's partition consists of two cells: P_A = {(U, L; U, R); (D, L; D, R)}. She distinguishes the first cell (where she plays U) from the second (where she plays D), but within each cell, she does not know whether Ben plays L or R.[39]

- Thus the actual current state of the world ω_c may not be discernible from other possibilities. If ω_c is that Ann plays D and Ben plays L, Ann cannot distinguish ω_c from alternate state ω_a in which she plays D but Ben plays R.

- A possibility operator represents an agent's conception of the world for a given actual state ω_c. Ann's *possibility operator* $P_A\omega_c$ = cell (D, L; D, R) shows the states that she considers possible when the actual state of the world is ω_c (where Ann plays D and Ben

[38] This game appears in Gintis (2009b, 41–44). He offers another case where neither player knows the other's γ and both use Bayesian probabilities for best responses to γ.
[39] Knowledge partitions are similar to information sets in extensive-form games.

plays L). In this circumstance, Ann knows she plays D and considers it possible that Ben plays either L or R.

- For any state ω, each agent i forms a *subjective prior* $\rho_i(\omega)$ that specifies her conjectures about the strategies used by others in ω. Technically, $\rho_i(\omega)$ indicates positive probabilities for all strategies of other agents that are considered possible in state ω. Having played D, Ann might conjecture that Bob plays L with probability 25% and R with probability 75%.

The *knowledge structure* of an epistemic game includes these three elements: knowledge partitions (P_i), possibility operators ($P_i\omega$), and conjectures about others' strategies for specific states ($\rho_i(\omega)$).

Combining concepts, an epistemic game proceeds as follows: Nature or an event sends a signal γ. On this basis, agents share an initial conjecture, with greater or lesser detail. They share a common element of their subjective priors $\rho_i(\omega; \gamma)$. In a minimally informed case, agents know only that everyone follows some suggestion from γ. By contrast, in a fully correlated case, equivalent to a focal point, agents know (or accurately expect) γ's strategy suggestions for the others.[40] Intermediate cases with various degrees of correlation are possible. Agents then act upon their conjectures: they play G on the basis of priors $\rho_i(\omega; \gamma)$.

The concept of *social epistemology* addresses how institutions (notably, conventions and norms) provide motivation, information, and cognitive frameworks for agents within epistemic games (Gintis 2009). As shared mental models, institutions exert precisely these influences: they offer motivational, informational, and cognitive context within which individual agents understand and make strategic decisions. In other words, institutions operate as correlating devices. Institutional prescriptions provide signals (γ) to agents, indicating correlated equilibria and focal points. Thus institutions are *social choreographers*; they act as social correlating

[40] Gintis' (2009b) knowledge structure also includes a knowledge operator that specifies the conditions under which specific events are known.

devices.[41] More specifically, institutions relate to epistemic game theory in three basic fashions. First, commonly understood institutional prescriptions shape agents' motivation via signaled social and material incentives. Second, and at a deeper level, normative prescriptions affect preference orderings, creating internalized values (ethical deontics) that promote adherence. Third, institutional signals convey information that influences conjectures about others' behavior ($\rho_i(\omega_c)$) in specific social contexts. Such common priors render others' behavior more predictable. Before meeting a stranger, I expect to shake hands. Before entering a classroom for the first time, students expect others to sit down at desks, as opposed to, say, dancing.

More fundamentally, as shared mental models, institutions indicate social categories—important elements of knowledge partitions (P_i)—and causal relationships that enter possibility operators ($P_i\omega$). A norm that prescribes precisely who (young people) should offer bus seats to elderly passengers partitions the social world of the bus into specific age-based categories. Likewise, the convention of driving on the right-hand side of the road (in the US) creates expectations concerning the actions of other drivers.

Overall, institutions perform *social choreography*: they orchestrate or correlate beliefs and behaviors across multiple individuals. Institutional signals—as specific, context-dependent ADICO prescriptions—facilitate strategic decision making by influencing and indeed shaping agents' motivation, information, and cognitive frameworks. Institutions thus provide and constitute shared mental models that guide social coordination; institutions indicate correlated equilibria and focal points. Such social choreography allows boundedly rational agents in non-ergodic environments to conceptualize alternatives, envision possible outcomes, and make strategic decisions. On such institutional foundations, agents navigate situations that would

[41] For Gintis, norms interpret events and motivate certain responses; they indicate correlated equilibria. Formal institutions share choreographic functions. Chapter 10 elaborates.

otherwise defy comprehension, on account of a multiplicity of possible actions, strategies, and equilibria. Institutions reduce the competence-difficulty (CD) gap enough so that agents can sometimes effectively offer best responses to the behavior of others in specific social contexts. Institutions thus enable boundedly rational agents to behave as substantively rational actors within certain contexts: as social choreographers, institutions create the potential for relatively predictable and consistent goal-oriented behavior among multiple agents—in certain contexts.