The Subjective Game Form
and Institutional Evolution as Punctuated Equilibrium

Masahiko Aoki
Stanford University

Summary: The major purpose of this lecture is to provide a conceptual framework for analyzing the mechanism of institutional evolution as “punctuated equilibrium.” Section 1 introduces reasons why an equilibrium view of institutions may be appropriate. Section 2 introduces a simple apparatus, called the COSE box, to represent the structure of the game. Section 3 then provides an equilibrium-related conception of institutions satisfying the four observed features: summary representation, continuity, multiplicity, and common knowledge. Section 4 drops the assumption of fixedness of agents choices sets in classical and evolutionary game theories and introduces the notion of subjective game forms. By discussing how the agents revise own subjective game forms in response to external shocks in a correlated manner, it attempts to describe a possible mechanism of institutional change. Section 5 deals with two important objective mechanisms for selecting new institutions: institutional complementarity and embeddedness. Section 6 provides some examples.

1. Why an Equilibrium Approach to Institutions?
Where do institutions come from? What role do they play? How do they inter-relate with each other within each economy and across economies? Can they be designed to change? Needless to say, these questions are impossible to answer without making explicit what is meant by institutions. However, leaving aside the old school of institutional economics, economists had largely been content with leaving the definition of institutions more or less vague. It is only recently that a few of them have taken up the task of conceptualizing institutions. Durkheim, a pioneer of modern sociology, once defined the discipline of sociology as the “science of institutions,” so such a task may have been thought to be outside the realm of

---

1This closing address at the Second World Congress of the International Society for New Institutional Economics, held in Paris on September 17-19, 1998, is based on Part II of my forthcoming book, Towards a Comparative Institutional Analysis (to be published by MIT Press). I benefitted greatly from research assistance by Christopher Kingston at Stanford University.
A technical formulation of the "rules-of-the-game" view has been presented by Hurwicz (1993, 1996). However, economists can make unique contributions to understanding the nature, origin, roles and consequences of institutions, but we must first make clear what we are talking about. There are at least three different (although inter-related) meanings that economists attach to the word “institution”. What we should be concerned with is obviously not a semantic clarification of the word as such, but a conceptualization that may be conducive to a better understanding of the workings of diverse economic systems by facilitating analytical insight.

In order to clarify the differences among the three meanings, or conceptualizations, of institutions that economists use, an analogy of the economic process with a game is apt. The game analytic apparatuses useful for institutional analysis, particularly those borrowed from the theory of evolutionary and repeated games, are of relatively recent vintage. However, the analogy of the economic process with a game can be dated back as far as Adam Smith, who stated:

“In the great chessboard of human society, every single piece has a principle of motion of its own, altogether different from that which the legislature might to choose to impress upon them.”
(1775:Part 6, Ch.2)

In the analogy of the economic process with a game, economists have regarded an institution as comparable to either a player of the game, the rules of the game, or an outcome (equilibrium) of the game. When people casually talk about institutions in daily conversations, they normally mean (prominent) organizational establishments, such as the government, universities, corporations, foundations, religious organizations, etc. Some economists follow this convention, effectively identifying an institution with a specific player of the game. However, North argues for a second view: that institutions should be identified with the rules of the game as distinct from players of the game. There are formal and informal rules of the game. By definition, the formal rules of the economic game cannot be constructed (changed) by the players of the game while they are playing, but need to have been determined prior to the playing of the game. As we are concerned with the origin of institutions, a question then immediately arises: Who determines the economic rules? It is here that North draws a sharp distinction between the rules of the game and the players of the game (organizations and their political entrepreneurs) who can act as agents of institutional change, i.e., as rule-makers. According to North, the existing rules of the game shape the incentives of the players (organizations) as to how to transact and what to innovate, ultimately generating effective demands for new rules in response to changing relative prices. The new rules will then be negotiated and determined in the

---

2 A technical formulation of the "rules-of-the-game" view has been presented by Hurwicz (1993, 1996).
“political market” that is structured according to political rules. North claims "[i]t is the polity that specifies and enforces the economic rules of the game."(North 1995: 22) Informal norms of behavior come from socially transmitted information and thus constitute a part of the cultural heritage. Therefore the process of their formation and erosion must be slow and complex. The tension between politically-determined formal rules and persistent informal constraints then may have important implications for the way economies change.

An equilibrium-theoretic approach to institutions was pioneered by Schotter and has been recently developed by Greif (1989, 1993, 1998), Milgrom, North, and Weingast (1991), Greif, Milgrom and Weingast (1994), Weingast (1997), and Young (1998) among others, who rely on sophisticated concepts of equilibrium in repeated prisoner’s dilemma games [such as subgame perfect equilibrium, and a variant of sequential equilibrium] or long-run stochastically stable evolutionary equilibrium.

I also subscribe to the outcome-of-the-game view (which we may also call the equilibrium view) of institutions. Depending on contexts, we may employ either evolutionary or repeated game approaches, and use correspondingly different equilibrium notions (some kinds of evolutionary equilibrium versus perfect equilibrium). Both approaches adopt rather extreme assumptions, in fact two polar opposite ones, regarding individual agents’ rationality in analyzing their motivations and choices. This may be interpreted as reflecting the present state of game theory, which has not yet succeeded in producing satisfactory models of the bounded-rationality of individuals, i.e., the trait of economic actors who are “intendedly rational, but only limitedly so”(Simon, 1961,p.xxiv). Thus we may regard these two approaches as being not fundamentally opposed to each other, but complementary. Regardless of which equilibrium notion we choose, the ones we employ are all Nash. That is, an institution is a socially constructed state from which agents are not motivated to depart as long as others do not do so. Also, whatever equilibrium notion is mobilized, it should be regarded as a stationary state of a process of interactions among the agents who actually play the game repeatedly over time, but not as a result of either perfect induction or completely passive inertia. Thus an institution as an equilibrium is “the product of long term experiences of a society of boundedly rational and retrospective individuals”(Kreps).

I intend to argue later that there is one important premise common to the evolutionary and classical game theories, i.e., the assumption of the fixedness of choice sets for players, that may set a limit in adequately dealing with the problem of institutional diversity and change. However, before delving into this issue, let me quote four reasons why the equilibrium-based approach may be useful for the analysis of institutions.

First, the equilibrium approach can deal with the issues of enforcement and of the origins of
institutions endogenously and in an unified way. If one subscribes to the rules-of-the-game view, then one must immediately face the issues of where and how the rules originate, as well as how the rules are enforced. Institutional origin may need to be found outside the domain of the economy in which the rules are applied: e.g., in the polity outside the economic domain or, theoretically, in a meta game in which rational agents collectively choose a set of rules from many such possible sets. But how, then, are the rules of the game in the political-exchange domain (polity) determined? How are the rules of the meta-game determined in which, presumably, all the technologically feasible strategic choices are provided? Thus, the problem of infinite regression seems bound to arise. It seems to me that the right way to solve this problem is to regard an institution as originating as an endogenous solution of a game -- closely connected to the Hayekian notion of cosmos meaning the spontaneous order -- in either the economic, social, or political exchange domain.

Behavioral standards or shared beliefs that emerged endogenously may eventually be articulated and codified to save various disequilibrium costs caused by mistakes, deviations, ignorance, etc. However, if such codification is not consistent with an equilibrium choice by the agents, the formalized rules – or what Hayek called thesis, meaning a “made law” or “‘set’ law” – will not be effectively enforceable or implementable. Of course, a game in which an equilibrium, and accordingly an institution, emerges, has by itself a certain structure. The same question may be then asked as to how these structures are determined. We can simply regard them as historically given. The development of institutions may then be regarded as a process of ramifications and refinements of games.

A similar problem of infinite regression can arise with respect to enforcement in the rules-of-the-game approach. If the rules of the game are to be enforced by augmenting a new player (enforcer), the question of how the enforcer is motivated to enforce the rules needs to be addressed. Do we need still another enforcer to enforce the standards of behavior prescribed for the original enforcer? As Hurwicz’s contribution suggests, a solution to this problem is again to analyze a game including the enforcer, if any, as a player, and see if the prescribed standards of behavior for the enforcer can become his/her Nash equilibrium strategic choice, given an equilibrium constellation of choices by other agents and vice versa. In other words, the equilibrium approach regards an institution as a set of self-enforcing constraints on an agent’s choices.

Second, the equilibrium approach to institutions provides an appropriate framework for analyzing interdependencies of institutions operating within the economy. When businessmen design organizational forms with the purpose of emulating better practices abroad, or when the government designs an organizational plan or drafts a statutory law for the purpose of introducing a so-far non-existent “institution” (such as markets for corporate control in the transitional economy), its implementation in particular economic, political, and social contexts can often have unintended consequences (such as insider control of
privatized ex state-owned enterprises). This is analogous to the situation in which a medicine which has been tested in a laboratory may have unpredicted side effects when it is administered to a human being due to the complexity of the living organic system. A major reason for such unintended outcomes could be the absence of “fits” between the designed plans and extant institutions. This suggests the possibility that only institutional arrangements that are mutually consistent and/or reinforcing may be viable in the economy. We can conceptualize such ideas as institutional embeddedness or institutional complementarities. As we will illustrate later, these intuitively appealing notions can become amenable to rigorous analysis when the equilibrium notion of institutions is applied.

Specifically, we can consider games in different domains of the economy, such as those for production coordination, commodity exchanges, transactions of human asset services, transactions of financial asset services, political-transactions, social-exchange, etc., and can analyze how an equilibrium constellation of strategic choices of agents in one domain can become strategically complementary to or conditional on the choices of the same, or different, agents in another domain. In this way, we can understand the conditional robustness of institutional arrangements. Also, understanding the structure of such interdependencies among institutions seems to be an indispensable prerequisite for developing a reasonable theory of the mechanism of institutional change.

Third, by showing the possibility of multiple equilibria in specific models, the equilibrium approach is able to focus on the “humanly devised” (North) nature of institutions rather than its ecologically, technologically or culturally driven aspects. If there is only one equilibrium corresponding to the technological specification of the structure of the game, then that equilibrium is little more than a disguised technological condition, but not an institution. For example, often the evolution of the community norms in East Asia is attributed to the climatic and ecological conditions there, which presumably make peasant family farming and collective use of the irrigation system more productive. However, Korea and Japan, which are characterized by similar ecological conditions, had rather divergent institutional paths in terms of village social structure and social norms, which may have had profound and long-lasting impacts on the subsequent differential institutional trajectories of both economies. Usually, a multiplicity of equilibria bothers game theorists, and they have spent much research effort on the so-called “refinement” of equilibrium, namely the refinement of the equilibrium concept to enable game theorists to identify only one equilibrium out of many possible Nash equilibria. However, we consider that the multiplicity of equilibria of games should not be regarded as bothersome in Comparative Institutional Analysis for the reason described above. We only need to carefully utilize empirical, comparative and historical information to identify important historical, political, and social factors that selected one equilibrium over the others in each economy. For an excellent exposition
of such methodology, see Greif [chapter 2, 1998].

Fourth, the equilibrium approach to institutions may also clarify the multi-faced roles of institutions. An institution is often conceptualized as consisting of self-enforcing constraints on the action choices of the agents. But in the world of incomplete and asymmetric information, an institution may also “enable” the bounded-rational agents to economize on the information processing needed for decision-making. Here, again, an analogy with the price mechanism familiar to economists may be somewhat useful. In the market mechanism, individuals do not need to know every detail of the economic environments in which they make their choices, but only relative prices (Hayek). Leaving aside the problem of the enforcement of contracts and property rights, if there were a complete set of markets, relative prices may be regarded as “sufficient statistics” summarizing the data needed for the society to achieve the social optimum in the most efficient way. Its dimensionality does not exceed the number of goods. (Koopmans, Hurwicz). Needless to say, in actuality there does not exist a complete set of markets. Individual agents therefore need alternative means to gain information useful in making their choices. Various institutions other than markets then evolve in response to the failure of complete markets to exist.

Just as markets transmit information regarding the economic environment (technologies, tastes and resource endowments) in the summary form of equilibrium relative prices, so do other institutions in alternative summary forms. I will later formalize this idea on the basis of the equilibrium notion. Although these forms are constructed socially and are thus internal to the economic system, they are nonetheless objectified as stable, “taken-for-granted” social facts by the agents precisely because they are derivatives of equilibria emerging out of the interactions of many. Individual agents engage in choice, taking as given the particular parametric values of these forms.

Just as there can only exist an incomplete set of markets, the ability of any institution to transmit information regarding the changing environment and choices of other agents in a summary form is also incomplete. But for individual agents who are bounded in their ability to process information and compute their optimal choices, such incomplete information may be adequate for making reasonably satisfactory choices in a relatively stable environment. However, such adequacy may become subjectively problematic when there is a drastic environmental change and crisis, or when a path of continual change crosses a certain threshold. Individual agents may then perceive that the taken-for-grantedness of institutional arrangements may not be tenable and begin to search for a new pattern of choice based on the collection of information, learning, experimentation, and so on, as well as the messages brought by existing institutions. As an aggregate outcome of such individual searching, agents’ expectations and representational perceptions about the internal and external state of the economy (or its sub-domain) may gradually converge and a new
institution may thus evolve. The bounded rationality of the agent can be a driving force of institutional innovation. Later, I will present a heuristic model to describe the mental models of the bounded-rational agents and discuss the mechanism of institutional change as a mutually equilibrating process of revision and refinement of such models by individual agents.

2. The Rules of the Game and the COSE Box

The primary object of my lecture is to build a conceptual framework for analyzing institutional change. However, as a preparation for that, I would like to start out with making precise my own conception of institution from an equilibrium perspective. Let us first take a particular sub-domain of the economy in isolation. The data which characterize the domain are

- the set of a finite number of agents (players), denoted by \( N = \{1, 2, \ldots, n\} \)
- the technologically feasible set (not constrained by human device) of choices for each agent, denoted by \( A_i \) with \( A = \times_i A_i \)
- the technologically feasible set of outcomes (consequences) of agents’ choice combinations, denoted by \( O \),
- the environment that is not under the control of the agents (natural, technological, and external institutional – i.e., institutions in other domains), but affects the physical outcome of agents’ choices, denoted by \( E = \{e\} \), and
- an outcome function that, given a state of the environment, translates each technologically feasible profile of choices by all the agents into a physical outcome relevant to their welfare, denoted by \( \phi: A \times E \rightarrow O \).

These data define the technologically (environmentally) determined “rules of the game.”

Suppose that given these rules, all the agents have expectations regarding others’ choices without specifying, for the moment, how such expectations are formed, and make a choice from their own choice sets according to respective choice rules. Thus, the endogenous variables are:

- an expectation of agent \( i = "_i \in A_i \) (\( i \in N \)); and
- a choice rule of each agent that, given an expectation and the knowledge of the outcome function, determines a choice from the choice set: \( \chi_i: A_i \times \phi \rightarrow A_i \) (\( i \in N \))
Then the structure of the game-form may be represented for any agent by the following 2×2 matrix, which we will refer to as the COSE box (An appropriate version when there are infinitely many agents, etc., may be drawn analogously). The left column represents the date to an agent, while the right to his variables. The first row refers to the dimension internal to the choice agent, while the second to that external to it. Each entry is self-explanatory.

### Box 1. Basic Game-form Structure for an Individual Agent (i ∈ N)

<table>
<thead>
<tr>
<th></th>
<th>parametric data (&quot;exogenous&quot; rules of the game)</th>
<th>endogenous variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>internal</td>
<td>(C) choice set</td>
<td>(S) strategic choice</td>
</tr>
<tr>
<td>external</td>
<td>(O) outcome function</td>
<td>(E) expectation of others’ choices</td>
</tr>
</tbody>
</table>

When the expectations of all the agents are consistent with actual choices of other agents, and if the choice rule for each agent is set by the best response rule with respect to a payoff function $u_i$ defined on the outcome space, i.e., if

$$
\alpha_{-i} = s_{-i}^* \quad \text{and} \quad s_i^* = \arg\max_{a_i \in A_i} u_i(\Phi(a_i, \alpha_{-i}, e)) \quad \text{for all } i \in N,
$$

the state is in *Nash equilibrium*. We have not specified how such consistency can be achieved. However, games which are relevant for the study of institutions must be those played repeatedly by the same agents, even if there are turnovers of agents over time. So we provide a steady state interpretation: “Each participant ‘knows the equilibrium and tests the optimality of his behavior given this knowledge, which he has acquired from his long experience.’”(Osborne and Rubinstein, p.5. See also Binmore)

There are normally many Nash equilibria for a game. In order for us to be able to predict certain properties of the outcome of repeated plays of a game that may be meaningful for institutional analysis, it is desirable for us to eliminate those equilibria that do not appear to be either relevant or meaningful. There are possibly two steps to doing it for institutional analysis: the first is to reduce the possible number of equilibria by theoretical reasoning, and the second is to mobilize historical and comparative information from outside the game that may be relevant to specific contexts of the game, and examine their possible implications for equilibrium selection from the many possible. For the first step, the notion of subgame perfect equilibrium (or
related sequential equilibrium) or some kind of evolutionary equilibrium concepts are normally employed. As these concepts are well known, I only display the COSE boxes corresponding to these notions below without further ado.

**Box 2. Sub-Game Prefect Equilibrium**

| (C) complete knowledge of feasible future choices | (S) deliberate, comprehensive contingent choice plan |
| (O) complete knowledge | (E) common beliefs regarding others’ choices off the play path |

**Box 3. Evolutionary Equilibrium**

| (C) fixed | (S) subjective pay-off improving choice with inertia (plus mutation) |
| (O) incomplete knowledge constructed from experiences | (E) inferences by limited memory |

3. An Equilibrium-based Conceptualization of Institutions

Now let us proceed to develop a conception of institutions on the basis of the game-form framework introduced as above. Although it is derived from the notion of equilibrium of a game, the proposed conception focuses on substantive characteristics of the equilibrium state perceived by the agents in common as relevant to their choice. Below I leave a concept of equilibrium unspecified, however. It may be loosely understood as an outcome of some kind of learning process which is robust with respect to random mistakes, experiments and deviants within a reasonable bound (thus sharing the characteristic of evolutionary equilibrium or subgame perfect equilibrium).

The following conception of institutions is based on four observations. First, in making their own choices, individual agents cannot form, and need not form, expectations regarding every detail of the choices that other individual agents will make. Agents perceive only limited, relevant aspects of the choices of others, or their intended choices in some important contingencies, on which basis they make a choice. It is intuitive that institutions *summarily represent* some essential aspects of the state of the economy. Second, a property
associated with a common conception of the “institution” is its durability or continuity, in that it is robust with respect to the continually changing environment, although this property may be subjected to a test when accumulated gradual environmental changes go beyond a certain threshold point. Needless to say, drastic environmental changes can also make an institution unsustainable. Thus, an institution cannot simply be identified with a particular (equilibrium) state of the sub-economy under consideration at a particular moment in time. It should be identified with something invariant within a certain boundary of environmental change.

Third, institutions are humanly-made orders. As such an institution is not a natural order which is uniquely determined by the technological and ecological environment of the domain of the economy under consideration. There can be multiple ways of establishing institutions under the same technological and ecological environment, if not for the same institutional environment. Fourth, there is a common understanding, a shared perception, among the agents in a particular domain of the economy about what constitutes an institution as relevant to their choices. If some humanly made construct is regarded as relevant and constraining their choice by some agents, but ignored as irrelevant by the others, it should not be recognized as an institution (although it may be among the former group of agents). So we seek a concept of an institution, derived from an equilibrium notion, but having the above four properties: summary representation, continuity, multiplicity, and commonality.

Suppose for an environment, $e \in E$, there exist multiple equilibria and that with each equilibrium $s^*(e)$ there is associated a function $E_i^*(\cdot)$ for each $i \in N$ that maps $A$ into a space $B_i$ with a smaller dimensionality such that

$$\text{given any } s \in A \text{ for which } E_i^*(s)=E_i^*(s^*(e)), \chi(s_i;\Phi) = s_i^*(e).$$

This implies that the value $E_i^*(s^*(e))$ sufficiently summarizes information regarding the equilibrium state $s^*(e)$ for the agent $i$ to make the corresponding equilibrium choice. That is, if he receives the same summary information as this in other (non-equilibrium) states, he would make the same choice as the one at $s^*(e)$ regardless of further details of those states. Next, let us assume that this summary representation procedure is continuous on a (connected) subset $S^\wedge$ of equilibria containing $s^*(e)$. That is, the nature of the information summary is qualitatively robust, provided that the environment changes continually so that the corresponding equilibrium does too. Formally, the function $E_i^*(\cdot)$ may be referred to as the i’s summary equilibrium representation. Suppose that for different equilibrium paths, $\{s^*(e)\}$, $\{s^{**}(e)\}$, ..., on $S^\wedge$ in response to a continual change in the environment, there exist corresponding continuous summary equilibrium
representations, such as $E_j*(\cdot), E_j**(\cdot), \ldots$. Finally, take the intersection (common elements) of the individual equilibrium summary representations over all the agents and define the collective summary equilibrium representation:

$$E^*_{s^*(e)} = \cap_j E_j^*_{s^*(e)}.$$ 

As each $E_j^*_{\cdot}$ is continuous on $\mathcal{S}^\wedge$, so is $E^*_{\cdot}$. We can do the same thing for $E_j^{**}_{\cdot}$’s, etc., to derive other collective summary equilibrium representations, $E^{**}_{\cdot}$, etc. We can readily see that each of these collective summary equilibrium representations satisfies all the four requirements mentioned above. It is somewhat imprecise though, therefore, we refer by $E^*_{\cdot}$, $E^{**}_{\cdot}$, ...etc., to the generic characteristics of a family of collective equilibrium summary representations, $\{E^*_{s^*(e)}\}_{s^* \in \mathcal{S}^\wedge}$, $\{E^{**}_{s^*(e)}\}_{s^* \in \mathcal{S}^\wedge}$, etc., that are invariant under any change in the environments as long as equilibria remain in $\mathcal{S}^\wedge$, and call them institutions. We refer to $E^*_{s^*(e)}$ as the phenomenon of the institution $E^*_{\cdot}$ at $s^*(e)$. Sequences of equilibria, $\{s^*(e)\}, \{s^{**}(e)\}, \ldots$ and corresponding institutions, $E^*_{\cdot}, E^{**}_{\cdot}, \ldots$ evolve in tandem and mutually reinforce each other.

Let us discuss the social contents of the functional form $E^*_{\cdot}$. It is the shared perception about the substantive nature of an equilibria path of the domain: the substantive nature robust to a continual change in the environment within a certain boundary, once established. As one of multiple possibilities for a given technological environment, it is socially constructed. Cognitively speaking, however, it represents the social reality that exists as if it is independent of individual agents and impacts on the outcome of their individual choices (Berger and Luckmann). At each state $s^* \in \mathcal{A}$ of the domain each individual agent takes the phenomenon of the institution $E_j^*_{s^*}$ as given and only need to process the residual representation information $E_j^*_{s^*}(s) - E^*_{s^*}(s)$ to make a choice $s_j^*$. Then he evaluates the outcome and pay-off of his choice by the simpler functions with the reduced domain: $N(a, E_j^*_{s^*}(s) - E^*_{s^*}(s))$ and $u_i(N(a, E_j^*_{s^*}(s) - E^*_{s^*}(s)) \in \mathbb{N})$. As the aggregate result of such individual choices, the equilibrium $s^*$ is generated, which in turn reproduce and reconfirm the institution. The constraint implied by an institution is self-enforcing in that it is the shared perception of agents about how the society operates (how the game is played) and it does not avail them to behave by ignoring it. At the same time, it enables every agent to make a bounded-rational choice without processing information about all technologically possible choices of others and their joint outcomes.

Thus we have the following verbal conceptualization of an institution corresponding to the technical
An institution is a stable, substantive characteristic of socially-constructed states of a sub-
economy which constrains agent's action choices through the convergent expectations it
generates, while enabling them to economize on information processing, provided that there can
be another such constraint for the same class of environments.

Some examples of the function $\Sigma^*(.)$ are:

- **Proto-property rights**: “One can catch as many beavers as one’s family has been doing for many
  years, but catching more than that could be troublesome either because beavers cease to reproduce,
  others punish by violence, or one simply feel bad.” (Demsetz)

- **Community norms**: “If one shirks in the production and maintenance of the village irrigation, he may
  be ostracized from social interactions in the village except for the time of fire and death of family
  members,” (Aoki) or “anybody is not expected to clean the drainage in the common.”;

- **Labor Market Institutions**: “If one shirk effort with the current employer, he will lose the well-paid
  job and have a difficulty in finding another job” (efficiency wage) or “if one shirks effort with the
  current employer, she will be able to easily replace him with another worker to make a good profit.”
  (Walrasian discipline);

- **Money**: “Money buys goods and goods buy money: but goods do not buy goods, as far as the price
  level is stable.” (Clower);

- **Organizational Conventions and Job Markets**: “If one quits the job with the current firm, he would
  have a difficulty in finding another job, as firms employ only new graduates out of school” or “if one
  does not pay the going rate of wage, then the workers will instantly move to other firms.”

- **States**: “If the sovereign tries to raise the tax without the consent of the council, his position would
  be imperiled” (Weingast) or “even if we try to resist the attempt of the village government official to
  raise the crop tax, it is not effective as he has a very strong backing of the Party” (predatory state).

As illustrated, most of institutions have the “if ...., then ....” form. In that sense, institutions become
a kind of “rules” (inference rules). They are the rules that are endogenously generated through the repeated
plays of the game. But, once established, they may be perceived as autonomous rules independent of
individual control, guiding and constraining their choices.

We may refer to the process of convergent expectation generated by evolving equilibrium as
institutionalization. It may take the form of spontaneous ordering — cosmos, in the terminology of Hayek
— such as norms, conventions, or cultural beliefs, or it may eventually be deduced as explicit rules —
nomos, also according to Hayek. We may call this type of institutionalization primary in that
institutionalization autonomously occurs within the original game-form structure. When institutionalization
is accomplished as a consequence of a change in the game-form structure, such as by adding an agent (an organization) equipped with a unique set of action choices (such as legal punishment), we may refer to it as secondary or derivative institutionalization — what Hayek refers to as taxis (‘made-’ or “arranged-order” in Greek). The emergence of the state as an enforcer of codified rules in place of common laws or private laws is an obvious example. In this case, the outcome function ought to contain information regarding the codified rules. The rules of the game now becomes partly socially constructed rather than completely technologically determined. However, mere codification of rules is not sufficient even for secondary institutionalization. The enforcement behavior of the enforcer needs to become predictable so that it becomes a convergent expectation that agents either observe the rules or are credibly punished. The important distinction here is that secondary institutionalization is also an equilibrium outcome of the game, while “set” law (thesis) sets a parameter of the rules of the game (an outcome function) that induces the equilibrium. As we will discuss below, institutional change or innovation often occurs as a consequence of structural change of the game-form relative to a prevailing institution. Thus we observe a spiral development of institutionalization such that a secondary institution eventually comes to serve as a primary basis for a further institutional change and so on ad infinitum. In the secondary institutionalization, the summary perceptual representation of an equilibrium may include a description about the shared expectation regarding a choice of a made “organization” that becomes the endogenous “rules of the game.” Thus three concepts of institutions: the player, rules and outcome of the game may come to be synthesized.

The following COSE box summarizes the role of institutionalization from this equilibrium perspective. I omit explanations.

**Box 4. Aspects of the Role of Institutionalization**

<table>
<thead>
<tr>
<th>An institution does........</th>
<th>through feedback to data</th>
<th>through internal feedback loops</th>
</tr>
</thead>
<tbody>
<tr>
<td>on micro (individual) dimension</td>
<td>(C) imputes values to accumulated abilities of the agent, defining incentives for capacity development (e.g., North)</td>
<td>(S) constrains/enables individual agent’s choice; habituates individual choice (e.g., Hodgson); regulates individual choice through <em>ethike</em> (moral sentiments) (e.g., Platteau).</td>
</tr>
<tr>
<td>on macro (collective) dimension</td>
<td>(O) coordinates collective adaptation to environments (e.g., Demsetz)</td>
<td>(E) stabilizes expectations on others’ behavior through <em>cosmos</em> (spontaneous order such as trust, norms, conventions, cultural beliefs) or <em>nomos</em> (derived rules)</td>
</tr>
</tbody>
</table>

\[
\begin{array}{llll}
\downarrow  \\
\text{micro-macro loop} \\
\downarrow  \\
\text{macro-micro loop} \\
\end{array}
\]
5. Subjective Game Forms and Punctuated Equilibrium

In the last decade or so, much progress has been made in understanding why institutions are robust and persistent in spite of environmental change. Factors often cited as deterring institutions from flexibly adapting to environmental change may include: specific code of communications (specific summary perceptual representation of an equilibrium, in our terms); various set up costs and increasing returns, complementarity and network externalities, incentive-institution reinforcing effects, etc. (see David for an excellent exposition of these factors). In spite of the possible roles of these factors, however, institutional change, marginal as well as in large scale, does occur. What is the mechanism for change? By the equilibrium-oriented definition of institutions, institutional change can be identified with a shift from one class of equilibria to another class such that a systematic, qualitative change in the strategic choices of agents, as well as their representational perceptions about the internal structure of the economy, occur. Initially, it may appear that there are two ways of realizing such equilibrium change. First, such a change may be thought of as occurring as a spontaneous ordering out of the decentralized experiments of agents in new strategies. Alternatively, equilibrium change may be thought of as introduced by a collective design of a new type of agents – organizations – equipped with a fundamentally different set of choices than the ones possessed by incumbent agents. The augmentation of such an agent will then change the information structure of the incumbent agents as well as the outcome of their choices, which may trigger changes in strategic choices on their side. Then, the key to an understanding of the second type of institutional change needs to be provided by clarifying how such a new organization is designed and implemented by the original agents. An orthodox theoretical approach to this analytical question can be to consider another public choice game, called a meta game, in which the original agents make a collective choice from the set of possible institutions (e.g., the legislative game in Reiter). But how is this political meta-game structured to begin with? As alluded to in the prologue of the book, there is a problem of infinite regression with this approach.³

Notwithstanding the apparent difference in substance, there is a common condition involved in bringing about a change by either route: either a critical mass of the original agents in the original game or a majority of the agents in a public choice game need to begin to modify their representational perceptions.

³Young [1998] constructs a game model of institution-selection based on evolutionary approach. In that an institution evolutionarily emerges and there is no deliberate collective choice, his approach may be thought of having affinity with the first type of institutional evolution. But the range of institutional choices is exogenous and fixed. Then, the long-run stochastic stability of the process leads to an optimal institutional arrangement.
about the internal state of the game, as well as its external environment, and jointly adopt new strategies in a
decentralized or coordinate manner. If we think this way, the distinction made above between the two
methods of institutional change may be fairly blurred. Even if there is ultimately a deliberate, collective
choice of the introduction of a new augmented agent (e.g., regulatory agency) to implement a change in the
agents’ choice, the accumulation of decentralized private experiments or a substantial agreement in thought
experiments across agents must precede it. But how do agents perceive a new opportunity for a change in
strategic choice and the need for a new organization to implement it, if necessary? Does it merely occur as a
chance event (mutation)? Alternatively, should the adoption of new strategies by individual agents be
regarded as a rational response to environmental change? If so, is institutional change uniquely and steadily
determined by the course of environmental change?

As a conceptual tool to deal with those issues, let us now introduce the subjective-game form, which
modifies the game form based on the classical (full rationality-cum-design) or evolutionary (random
mutation-cum-learning) approach. In this form, individual agents are assumed to have limited, subjective
perceptions about the structure of the game, accumulated from the past, and accordingly make bounded-
rational choices. Using this tool, we analyze how environmental change or accumulated internal pressure
affect the representational perception of an individual agent about the internal and external environment of
the game and its strategic choice in a non-deterministic way. Specifically, we depart from the classical or
evolutionary game approach in abandoning the assumption of the fixedness of choice sets of the agents.
Because of this assumption, evolutionary game theory predicts that the long-run stochastic evolutionary
equilibrium tends to be the efficient one in the set (Young 1998), and the classical game theory has to regard
institutional change as events occurring outside context-specific models.

Recall that in the basic game form structure as displayed in the COSE box 1, the entry (C) is
assumed to represent the set of all technologically feasible choices of an agent in a domain, while entry (O) is
assumed to summarize all technologically feasible outcomes contingent on some technologically feasible
choice profiles. We can incorporated the possibility in which the physical outcome of the game may be
affected by the equilibrium choice profiles in other domains. In any case the classical game theory assumes
that, given the rules of the game as represented in the left column of the COSE box, and, given an expectation
about other agents’ choices represented in entry (E), the individual agent is assumed to make the best
response choice. Now let us assume that the individual agent cannot have full knowledge of the
technologically determined rules of the game nor can he make perfect inferences about other agents’ strategic
choices. Instead we assume that each agent’s perception of the structure of the game is represented by the
following “subjective” game form.
• (C) The objective set of “technologically feasible” choices of the agent \(A_i (i \in \mathbb{N})\) may be represented in a space of infinite dimension, but only its finite-dimensional subset is activated for possible choice at any moment in time. A subset may be chosen and sustained once for a certain duration of time in a manner specified below. We call this subset the **activated set** of choices (technically it is a hyperplane of the technologically feasible choices) and represent it by \(A_i(t)\), where \(t\) denotes a parameter representing a finite combination of dimensions which span a subset of the technologically feasible choice set.

• (E) The agents common perceptual representation of the internal state of the domain is given by the phenomenon of an institution \(E^*(s^*)\) when the profile of strategic choices by agents in the domain is \(s^*\).

• (O) Actually the phenomenon of the institution contains the summary information concerning the state of the environment as well, because agents’ choices are partly responses to the environment (through their perception of the outcome function). However, we assume that the agent additionally processes information regarding a part of the environment (technological, international, or external institutional) of the domain that is relevant to him/her and forms an individual perceptual representation of the environment. When the true state of the environment is \(e \in E\), an agent’s perception about the environment is represented by his private estimate of probability distribution, \(\varepsilon_i (i \in \mathbb{N})\), on a partition of the space of the environment \(E\). These representations are imperfect in the sense that the partition of the space is coarse and (infinitely) many states of the environment are not distinguished from each other with the same (or zero) probability attached. Let \(P\) be the set of probability distributions over \(E\). We call the rule: \(\psi_i: E \rightarrow P\) which maps the true state of the environment to its private perceptual representation the agent’s environmental inference.

• (S) The agent makes a choice from his activated choice set that, given a particular phenomenon of the institution [and the residual individual summary representation of the internal state of the domain], is predicted to maximize his/her utility (pay-offs) according to his/her perception of the environment. Namely, the agent chooses \(s_i^*\) that maximizes \(u_i(a_i; E^*(s^*), Q_i(e))\) subject to \(a_i \in A_i(t)\), where \(u_i(.; E^*(.), Q_i(.)\) is the pay-off predictor. We call this operation the **best-response choice rule**.

Then we have the following COSE box representing the subjective-game in which each agent makes a choice, according to his or her subjective inference and prediction, within the constraint of the activated choice set.

**Box 5. Structure of Subjective Games**

| \(C\) \(A_i(t)\) = the activated sub-set of choices | \(S\) best-response choice rule: \(s_i^* = \arg\max_{a_i \in A_i(t)} u_i(a_i; E^*(s^*),.)\) |

---

16
When the agent repeatedly uses the same rules for environmental inference, pay-off prediction, and choice, as well as the same phenomenal perception of institutions, we say that his/her subjective game is reproduced (or in equilibrium). More generally and realistically, it may be considered that the agent possess multiple rules of inference and prediction at one time which are mutually competitive in some respects, but complementary in others. Then, given a continually changing institution, the agent may experiment with each of them and choose the one that he or she considers appropriate under a given circumstance. However, when the sets of multiple rules remain to be stored as useful tools by the agents, we may still say that their subjective games are reproduced. Note that the subjective games in this sense roughly correspond to the notion of “mental models’ in the induction theory of Holland and et al. They conceive of the mental models as “models of the problem space” that cognitive systems construct, and then “mentally ‘run’ or manipulate to produce expectations about the environment.”(Holland; 12). But our analysis differs from theirs in that we explicitly consider not only the interactions of such models with external environments, but also with those of other agents who act in similar ways (see also Denzau and North).

Suppose now that an institution $E^*(.)$ is historically given in a rule form as illustrated before. There, those rules were regarded as being generated and self-enforced as a joint outcome of choices of the agents who play an objective game. We can extend such a notion to the present case where the agents each play their respective subjective games simultaneously. When it holds that

$$s_i^* = \text{argmax}_{a_i \in \Omega(T)} u_i(a_i; E^*(s^*), Q_i(e))$$

for all $i \in \mathbb{N}$ and for all $e \in \mathbb{E}^\wedge$

we say the institution $E^*(.)$ is reproducible (the rules are self-enforcing) on the sub-class of possible environments, $\mathbb{E}^\wedge$. This is the situation in which all agents perceive the phenomenon of the institution and act accordingly, and as a result, the institution is reconfirmed and reproduced as a germane guide/constraint for agents’ choices. The reproducibility of the institution may not necessarily require the rigid reproduction of the subjective games played by the agents. The agents may marginally change their sets of rules for personal perception, pay-off prediction, and choice, but it is possible that the above “general equilibrium” condition could still hold for a given institution.

When his/her existing set of rules does not produce satisfactory results relative to an agent’s
aspiration, however, the agent may start revising/refining the existing set of rules more substantially, in particular generating new choice rules involving the expansion of the activated set of choices. But when is such a gap between aspiration and prediction, that is, subjective disequilibrium, likely to occur, especially in critical mass? An answer is that this may happen in the case of a drastic change in the environment, as a result of cumulative impacts of equilibrium sequences on the environment and pay-off distributions, or possibly by both. As external triggering conditions, we may think of the following events:

- new technological innovation occurs so that new choices become feasible (hitherto inactivated dimensions of the choice sets are invoked);
- external shocks, such as war, perceived productivity and innovation gaps with foreign competitors, prolonged depression, compel the agents to perceive a need for rapid improvement in productivity and other performance characteristics;
- a large scale institutional change occurs in a neighboring domain with which strong institutional complementarity exists.
- A large change in policy determined parameters of the outcome function occurs.

These cases may be represented as the environmental parameter moving out of a certain subset $E^*$ which warrant the equilibrium of the subjective games of the agents. When such an environmental change occurs, the extant activated sets of choices, as well as associated subjective rules – ones for inferences, predictions and choice – are re-examined, revised or replaced by ones that may be considered to “work”. A search for the re-definition of the subjective game can be initiated by an individual agent who perceives new opportunities even under a fairly stable environment. If proved successful, such a redefinition may be emulated by other agents and eventually become self-organizing as a spontaneous order. But, when the performance characteristics of the domain are satisfactory and no significant gap is perceived by the agents, the impact of such an entrepreneurial mutation may be limited (Aoki 1998). It is rather likely to be a general perception of large disequilibrium in their respective subjective games that initiates a synchronized search for the redefinition of subjective games among agents. Such a search is directed by problem-solving incentives and based on feedback of the success or failure of experiments the agents have had. The type of triggering condition obviates random search. Instead, the agents direct their search according to an emergent problem situation, generating rules that are likely to be useful for it and hence possibly useful in the future as well. Thus, search activity becomes highly context-specific.

Once simultaneous revision (innovation) by many agents of their activated choice sets and the implementation of new choices start, the hitherto existing institution as well will cease to provide a useful
guide/constraint for individual choice, because it will not be effective in summarily representing newly emergent choices. The “taken for granted” premises implied by the institution will be questioned. Agents now need to process a larger amount of information regarding the internal state of the domain than they did under the rules of the old institution. In particular, they have to process information regarding emergent choices that may be germane to their pay-offs. However, as these synchronized search activities are induced by a common triggering condition, their directions are not randomly distributed, but may be highly correlated. Another factor could be that agents draw analogies and seek alternative models from other domains (including those in foreign economies) and from their (shared) history. Finally, experimental new choices may prove to be more effective if there is some kind of fit (complementarities) with extant institutions or emergent new experiments in other domains. Or, they may be introduced as effectively coordinating strategic choices in hither-to separate domains so as to realize an innovative institutionalized link, entailing dis-bundling of old links. Using these and other possible reasons, there are likely to evolve at most a few major types of competing models of subjective games in response to the trigger. Competition among them characterizes the transitional process.

A new model of subjective game for an agent is equilibrated, when (1) an application of a new inference rule does not yield a big surprise; and (2) a choice from a new activate choice set is perceived to be generating a satisfactory outcome as predicted. The transition process comes to rest only when the continually revised models of the subjective games of agents become consistent with each other in the sense that they are simultaneously equilibrated. Then there will evolve a new way $\Sigma (.)$ of summarizing a new state of (moving) equilibrium, providing the agents with a basis for stable summary representations of the internal state and implicitly the new classes of environment. Thus, the actual process of institutional evolution may be characterized more as what biologists Stephen Jay Gould and Niles Eldredge conceptualized as punctuated equilibrium rather than the steady, gradual Darwinian selection process. An evolutionary process characterized by punctuated equilibrium is one in which long periods of stasis are broken by short, in geologic time, episodes of rapid speciation. Although biological metaphor and analogies cannot be perfect, nonetheless their concept is highly relevant and appropriate. Once a particular system (institutional or biological) is established, it tends to sustain itself. Change in the system may be more likely to be initiated by a large external shock rather than one that is slow and gradual. Characteristics selected during one point in time impose constraints on future possibilities (path dependence). On the other hand, however, it is not

---

4Danzau and North also refers to this analogy. As far as I know, one of the first social scientific references to this analogy is by Stephen Krasner [1988]. Fudenberg and Maskin (1992?), as well as Young (1998), constructed and analyzed evolutionary models having the characteristic of punctuated equilibrium.
certain whether the transition to the emergent institution was the only possible trajectory from the initial state of disequilibrium. Far from equilibria, branching out along multiple paths may be possible (Aoki 1988). Therefore, for an understanding of the mechanism of institutional change, careful empirical studies based on historical and comparative information are necessary to sort out what major factors are influencing the selection of a particular trajectory instead of other possible ones.

Box 6 summarizes the mechanism of institutional evolution. From the left it deals with the choice of endogenous variables in the “old” subjective game, its feedback to the data of the old subjective game and its redefinition, and the emergence of the “new” institution.

**Box 6. The Mechanism of Institutional Evolution**

<table>
<thead>
<tr>
<th>Solution under the Old Institution</th>
<th>Feedback to, and Redefinition of, Rules of Game</th>
<th>Evolution of “New” Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S) choice constrained by the existing activated set of choice $\rightarrow$</td>
<td>(C) gap between aspiration and achievement $\rightarrow$ context-specific search for a new model of the subjective game $\rightarrow$ redefinition of a new activated subset of choices $\rightarrow$</td>
<td>(S) novel strategic choice</td>
</tr>
<tr>
<td>(E) old institutions</td>
<td>(O) environmental change (technological change, external shocks, change in complementary institutions in neighboring domains)</td>
<td>(E) new institutions</td>
</tr>
</tbody>
</table>

6. **Diachronic Linkage of Institutions and Path Dependence**

In the previous section, we have discussed the mechanism of institutional evolution from the angle of individual subjective views of the rules of game and their interactions. In this section, we discuss an (objective) aspect of the mechanism of institutional evolution: how the nature and process of diachronic institutional linkage may affect the course of institutional evolution. In the previous section, we suggested that during the period of institutional change (at the time of “punctuation”), various experiments in agents’ choices (in response to external shocks) take place and compete for selection as viable alternatives to well-worn choices. There could be many sources or grounds for viable experiments. But the following two mechanisms seems to often play an important role in realizing an institutional innovation and change, or alternatively deterring such change. First, a type of newly activated choice may not be viable as a stand alone experiment, but if a complementary experiment or change takes place simultaneously, mutual reinforcement between the two may induce a persistent tendency toward institutional change. This is the mechanism of diachronic (intertemporal) complementarity. As we will see presently in one example, fits need not be well designed or expected *ex ante* by experimenting agents, but can be incidental and unintended.
Second, an institutional innovation may emerge as a new type of bundling or inter-mediation of hither-to-separated, old games, possibly entailing dis-bundling of older institutionalized links. By bundling, I mean the event in which some agents coordinate their strategies across games in different domains and make a strategy profile an equilibrium (i.e., an institution) in linked games which was not possible in isolated games. Indeed, according to Joseph Scumpeter, an economic definition of innovation is “a new combination of economic factors formed on the basis of creative destruction of older combinations,” and the entrepreneur is conceptualized as the agent for carrying out such creative destruction. An understanding of the first mechanism can shed light on how embryotic experiments can or cannot grow into an institution, while the second clarifies one of the important venues through which agents’ activated sets of choices can be expanded to create a possibility of new institution. In any case, if institutional change, creation, or sustenance is indeed impacted by such mechanisms, extant structures may delimit the range of possible options for institutional evolution. This suggests the highly path-dependent nature of institutional evolution. Hence a satisfactory analysis of the mechanism of institutional change could be made only in specific historical context.

(A) Diachronic Complementarity
Assume that there are two games $G1$ and $G2$ with the respective set of agents having respective homogenous pay-off prediction rules $u$ and $v$, and that there are only two classes of equilibria cum institutions in each game, denoted by $(E1*; E1**)$ and $(E2*; E2**)$. For simplicity’s sake, we assume a one sided complementarity: i.e., the payoff differences $\Delta E = u(E1*) - u(E1**)$ for the agents in $G1$ are greater if $E2*$ evolves in $G2$ rather than $E2**$. However, the pay-offs of the agents in $G2$ are not directly affected (or at least perceived that way) by choices of the agents in $G1$. We adopt a convention to order equilibria as $E1* > E1**$ and $\Sigma2* > \Sigma2**$ (it does not imply any substantive characterization).

We assume that there are parameters $2$ and $\eta$ characterizing the (subjective) game forms of the agents in $G1$ and $G2$ that affect their pay-off predictions. Any parameter can be taken, such as representing agents’ perceptions of the technological environment, their competence and stock of knowledge, law and regulations, government policy instruments that affect the outcome of the game, etc. It is assumed that these parameters are institution-sensitive in the following sense. They are ordered along the real line $\Re$, or take binary values, say 1 or 0 (in the case of $E$’s ordering, this ordering convention does not have any substantive meaning), and each of the pay-off prediction rules has increasing differences in its own variable and its relevant parameter, that is, $u(E1*, E2, 2) - u(E1**, E2, 2)$ is increasing in $\theta$ for either value of $E2$; and that $v(E2*, 0) - v(E2**, \eta)$ is increasing in $\eta$. These specifications imply that $2$ and $\eta$ are measured in such a way
that their higher values (alternatively lower values) have a relative (not necessarily absolute) fit with $E_1^*$ and $E_2^*$ (alternatively $E_1^{**}$ and $E_2^{**}$) respectively.

We need to specify how these parameter values can shift. We assume the following dynamic process under normal circumstances: denoting the values of the variables and parameters at time $t$ by postscript $(t)$, the states of the parameters are determined in such a way that $Z(t+1) = F(Z(t), 0(t), E_1(t), E_2(t))$, and $0(t+1) = G(Z(t), 0(t), E_1(t), E_2(t))$, where $F$ and $G$ are non-decreasing functions and strictly increasing with respect to their own parameters, i.e., there is no natural depreciation in the value of parameters). Roughly, this implies that the parameter values will not autonomously change in the direction opposite to the way the system as a whole is set. For example, if prevailing institutions are $E_1^{**}$ and $E_2^{**}$ and the parameter value of $0$ is low so that it relatively favors $E_2^{**}$ in $G_2$, then the parameter $\theta$ will not autonomously increase to make $E_1^*$ relatively more attractive in $G_1$, but possibly decrease. If $\theta$ measures the institution-sensitive competence of the agents, it means that its accumulation is in the direction consistent with the prevailing institution (say, through learning by doing), as well as with the prevailing complementary institution (through the direct complementarity effect of $E_2$ and the indirect effect of $0$, say, policy instruments adopted there to sustain that institution).

Finally, assume that the values of $E_1$ and $E_2$ are chosen in $G_1$ and $G_2$ in each period $t=1,2,\ldots$ so as to maximize the joint pay-offs, $u+v$. If the agents in $G_1$ and $G_2$ are overlapped, then it is equivalent to assume that the agents maximize the separable pay-offs, $u+v$. If not, the assumption may be rationalized when the outcome of an institution $E_2$ can be “sold” to an institution $E_1$ as an input, while the proceeds deliver payoffs to the agents in $G_2$.

With these set-ups, we can examine the effects of institutional complementarity on the mechanism of institutional evolution. Without a proof, we claim the following variant of the momentum theorem initially due to Milgrom, Qian, and Roberts.

**Proposition 1.** Even if the initial level of institution-sensitive competence is too low to make the institution $E_1^*$ viable in $G_1$, the presence of $E_2^{**}$ in $G_2$, together with a modicum of an initial increase in the institution-sensitive competence may induce its steady growth so as to make the institution $E_1^*$ eventually viable.

**Proposition 2.** Suppose that the institutions $E_1^{**}$ and $E_2^{**}$ have prevailed up to time $S$, at which time an exogenous parametric change (e.g., policy change) occurs in the (subjective) game form of $G_2$
to make the relative (not necessarily absolute) payoffs from an alternative institution $E^*2$ improve. Then, there can be endogenous spiral growth in the values of the structural parameters $\gamma$ and $\omega$ from then on, which will successively improve the relative (not necessarily absolute) payoffs from the institutional combination $E^*1$ and $E^*2$. If either the value of $\gamma$ or $\omega$ becomes sufficiently large in this process, $E^*1$ and $E^*2$ may be eventually institutionalized.

Thus, we see that the initial presence of a complementary institution, or a reversal of the direction of change in specification of (subjective) game forms, can trigger the emergence, or change, of an institution in another domain with which institutional complementary relationships may exist. The subtlety of the process can be that the emergence or change of an institution can be only latent at the outset and gradual, but that gradual change in parameters specifying (subjective) game forms in both domains plays a crucial role in the process of transition. Another important point is that, in spite of the assumption of joint maximization, fits between emergent institutions does not need to be intended or planned beforehand. In fact, as some of examples below show, institutional dynamics triggered by some parametric change, such as in policy instruments and propagated through the institutional environment, may have unintended consequences that policy makers had never thought of.

(B) Overlapping Embeddedness and Path Dependence

In (A), we dealt with the diachronic linkages of institutions generated by the agents’ interactions of strategies across different domains, aided with those of institution-sensitive parameters. In this sub-section, we deal with the diachronic linkages of games generated by the possibility of agents strategically coordinating their own choices across domains. As we did not exclude the possibility in the previous section that the agents may be overlapped across domains, the distinction between the two may be blurred. However, there are some interesting examples in this special case having sharp implications for path-dependence. Also they may illustrate some interesting cases for the emergence of a new type of strategic choice.

Suppose that in the game $G1$ an institution $E^*$ has prevailed. Imagine that, through an external contact, a new game $G2$ is linked to that game with all the agents in the older game included in the new game and thus embedded. Suppose that a new institution $G^{**}$ emerges for the linked game $G1 \cup G2$, generated by, and supporting, a new type of strategic choice by the original agents which was neither possible in $G1$ nor $G2$ in isolation. This can be regarded as the old institution $\Sigma^*$ evolving to the new one $\Sigma^{**}$, or the new institution $\Sigma^{**}$ emerging in $G2$, when they are embedded in a larger game. Given the multiplicity of institutions, the
nature of this new institution is likely to be conditioned by the old institution rather than solely determined by
the technological structure of the linked game. Suppose further that, by the increasing enhancement of the
domain of the game $G_2$ or for some other reason, the domain of $G_1$ gradually shrinks and finally disappears.
Correspondingly, a further new institution $\Sigma^{***}$ may emerge (or the institution $\Sigma^{**}$ may remain) in the game
$G_2$ (which may be linked to a newer game). However, this institution may differ from the one to prevail when
$G_2$ stands alone from the beginning, or is/was linked to a different game than $G_1$. Thus the historical path of
linkage may leave its imprints on the nature of succeeding institutions even after the structure of linked games
changes.

7. Examples

(A) Diachronic Institutional Complementarity

Example (i): The Silicon Valley Phenomena. The fascinating story of the Silicon Valley phenomena all
began at the place where a small number of risk-taking venture capitalists and a small number of
entrepreneurs started to meet. The venture capitalists started to experiment on a new strategy in the domain
of financial transaction – say, venture capital financing strategy characterized by a small amount of initial
start-up funding, followed by step-financing contingent on the smooth progress of projects, and culminating
in the market or private sales of the successful firms, whereas the entrepreneurs started to experiment on a
new strategy in the domain of organizational formation – say, “information encapsulation” strategy to pursue
independent innovation projects outside the context of traditional corporate hierarchies. As analyzed in detail
in chapter 12 of my forthcoming book, both strategies are potentially complementary, but for them to be
effective they need to be backed up by “institution-sensitive” competence. The venture capital strategy can
become more effective when venture capitalists have a risk-taking attitude more firmly based on management
know-how lacking in start-up entrepreneurs and the technological expertise necessary for judgement on step-
financing. The information encapsulation strategy can be more effective when start-up entrepreneurs have a
deeper as well as broader information processing capacity that would facilitate independent product design
innovation in niche markets. However, in the beginning there was not enough competence to institutionalize
both strategies.

But, several factors, some accidental to the region, started to make this competence accumulate: the
presence of a research university and other educational institutions in the neighborhood that had the culture of
more open intellectual communications and a supply of risk-taking graduates; the presence of skills in
machinery manufacturing, inherited partly from the war-time defense industry, that facilitated the fast
manufacturing of customized pilot products; the development of information and communication technology; government policy to deregulate the portfolio selection of pension funds in risky investments and to lower capital gains tax; and the ecological factors to make potentially tense intellectual life more enjoyable. Once all these elements are put to work together in one place, the initial accumulation of the germane competence got started. Learning from doing, as well as from partners’ activities, combined with inherent complementarity between the two strategies, provided momentum, further attracting the relevant resources from the outside. Both strategies became established as mutually supportive equilibria rather than an outlier and thus as an institution.

Example (ii): Unintended Fits in mid-twentieth century Japan. After the prolonged depression and in the preparation of the coming war, the era of the relatively laissez faire economic policy in Japan came to a close in the 1930s. The military-bureaucracy collusion tried to remove capitalist control of enterprises by limiting their legal rights in corporate governance and heavy property taxation, substitute government control over bank lending for competitive capital market allocation, organize an industrial control association in each industry as an intermediary for command transmission and information collection, order technological transfer by larger firms to smaller firms, and form all-inclusive employees organizations (called the Industrial Patriotic Society) at all factories and establishments. All these measures were adopted and coerced for strengthening the centralized control over the war economy. Some of measures mentioned above were indeed an emulation from the “success” of the economic planning in the U.S.S.R in the late 1930s (Okazaki). However, these measures were never came to be implemented as self-enforcing strategic choices in the private sector. With the disastrous failure of the centralized economy and defeat of the war, the military was removed from the government and the representative government was installed.

Under the direction of the Occupation Army, the Japanese government implemented various policy measures to democratize the centralized control over the economy. However, as Table 1 exhibits, the organizational and structural policy measures taken by war-time Japan continued to be adjusted in the democratic context and set initial parameter values specifying the structure of various domains of game in the post war economy. Those structural arrangements appear to have been highly functional in the period of high growth (1950s and 60s) and generated various unique institutions as exhibited in the third column (see Aoki 1988 chapter 3 for the enterprise unionism. For the contingent corporate governance and main bank system see Aoki 1994, 1997). Why were they capable of being successful in spite of the fact that they were mal-functional during the war? One possibility is that some change conducive to a productivity increase evolved in the private sector in and after the transition from the semi-centrally-controlled economy to the
democratized market economy, and that it found "unintended fits" (Aoki 1997) with the government designed structural arrangements, thus making them an “over-all spontaneous order.” I pinpoint such a change in the workplace of the firm.

During the war, severe shortages of man power, materials and tools on the shop floor required ad hoc cooperation of the workers for problem-solving (I discuss below how it became possible incentive-wise). The emergent practice of cooperation set the value of an incentive parameter at the shop floor to reward the worker more for developing “contextual skills” rather than “individualized skills.” Such strategy evolved into a convention in the process of high growth and the associated organizational convention of horizontal hierarchies was established as an institution then. In this process, institutions as summarized in the third column co-evolved in other domains as a complex of complementary institutions (see Aoki 1994 for the analysis of institutional complementarity). This case may be taken as an archetypical example of Proposition 2, where the initial reversal of parameter values helps generate unintended institutional evolution through a mutually reinforcing manner.

Table 1. The Structure of Unintended Fits

<table>
<thead>
<tr>
<th>domains</th>
<th>policy parameters set by the war-time government for central control</th>
<th>policy parameters set by the post-war reforms</th>
<th>institutions co-evolved with horizontal hierarchies in the high growth period</th>
</tr>
</thead>
<tbody>
<tr>
<td>corporate governance</td>
<td>limiting stockholder’s legal rights</td>
<td>zaibatsu dissolution</td>
<td>micro corporatism</td>
</tr>
<tr>
<td>financial transaction</td>
<td>command implementation through credit control</td>
<td>credit rationing for reconstruction</td>
<td>contingent corporate governance (chapter 10)</td>
</tr>
<tr>
<td>government-business relationship</td>
<td>industrial control associations</td>
<td>consultative industrial associations</td>
<td>the main bank system</td>
</tr>
<tr>
<td>worker’s organization</td>
<td>the Industrial Patriotic Society</td>
<td>legislation expanding workers’ rights</td>
<td>bureaupluralism</td>
</tr>
<tr>
<td>supply relationship</td>
<td>guiding cross-firm technological transfer</td>
<td>protection of small firms as social policy</td>
<td>multi-vendor subcontracting system</td>
</tr>
</tbody>
</table>

(B) Overlapping Embeddedness
Example (i): The Formation of Cultural Beliefs. The story told by Greif about the Maghrebi merchants may be liberally adapted as follows: the norm of ostracization in the event of inappropriate conduct was formed in the community game of the Maghrebi traders prior to their involvement in long-distance trade. When they initiated long-distance trade and the prospects of sustaining a reputation mechanism in this domain alone was still weak because of high time discount, high risk in trade profits or some other reason, the merchants might have conceivably coordinated their strategies in the trade game with those in the community game so that a deviant in the trade game is ostracized from participating in the sharing of social surpluses in the community game. As rents from honest trade became high enough, the sustenance of the reputation mechanism supported by the “collective cultural belief” became feasible, which differed, however, from the one operating among Genoese who had a different historical background.

Example (ii): Learning and Hybrids. One possible impact of diachronic embeddedness may be found in the context of learning from a foreign practice. When producers in one economy recognize a large productivity or innovation gap with foreign competitors, they may try to emulate the competitor’s practice which they perceive to be superior. Learning will certainly modify the structure of their models of subjective games, in particular by adding new action choices. But such expansion or refinement of the activated choice sets can be realized not only through their own perceptual representation of foreign practices, but also in the context of longer-lived embedding institutions, such as group norms at the work place or market-oriented ethics, etc. What is perceived as transplantable may be filtered in the context of embedding institutions. Then learning is unlikely to result in a simple transplant of the set of action rules from abroad. Rather it is more likely that the perceived set of new choice rules is combined with the indigenous set of rules to generate a hybrid. Such a hybrid may sometimes result in a degenerate model of subjective games, failing to achieve the original intention of learning. However, in some cases, a hybrid may activate a set of innovative choices hitherto unknown.

Example (iii): Innovations in Corporate Information Processing. The Silicon Valley phenomena emerged in the relative periphery of Industrial America. However, it is still within broader Industrial America. Meanwhile, the same phenomena has had a harder time emerging in other places. There can be many reasons for this, but we discuss one possible path-dependent impact.

As suggested, typical start-up entrepreneurial firms in Silicon Valley encapsulate information processing relevant to development effort from each other. It has two aspects: information processing is integrative to the degree necessary for the independent design of a modular sub-system, on one hand, while
being specialized to the degree geared toward the design of a unique and original modular sub-system, on the other. If we view the clustering of Silicon Valley firms as one large information system, then, the information processing of each unit -- i.e., each entrepreneurial firm – has more breadth relative to that of component task units in traditional integrated firms, while it has more depth relative to that of component task units of horizontal hierarchies tightly knit with internal information feedback mechanisms. The breadth of information processing may be gained by the holders of individualized, function-specific skills through the aid of improved information and communications technology. However, gaining its depth may be harder for the holders of contextual skills without sacrificing non-technology-based communications, which may require a fundamental change in some essential aspect of horizontal hierarchies. Thus the Silicon Valley phenomena may have evolved more easily, combined with other conditions as mentioned before, in the periphery of Industrial America where the accumulation of individualized, function-specific skills was relatively easier to recruit.