

## . THE INSTITUTIONAL FOUNDATIONS OF A MARKET ECONOMY WDR 2001/2

### What are Institutions? How Should We Approach Them?

Masahiko Aoki

Stanford University

#### *Aspects of Institutions: Shared Beliefs, Summary Equilibrium Representations, and Endogenous Rules of the Game*

From the viewpoint of an analogy of the economy with a game that can be dated back as far back as Adam Smith, there are three major conceptualizations of institutions:

Players of the game (government, courts, industrial associations, etc.) view: e.g., Nelson

Rules of the game view: North, Hurwicz, Ostrom.

Equilibrium of the Game view: Suggame Perfect (expectations), e.g., Greif, Greif, Milgrom and Weingast, Milgrom, North and Weingast, Calvert. Evolutionary Equilibrium, e.g., Sugden, Young, Bowles, Aoki.

In this presentation, I propose to develop a game-theoretic approach to institutions that is essentially in accord with the institution-as-the-equilibrium-of-the-game view, while incorporating some aspects of, the institution-as-the-rules-of-the-game view. I tentatively characterize it as an attempt to understand the nature of an institution as a *collectively-perceived, summary representation of an equilibrium* of the game (out of the many that are theoretically possible). That is, an institution is viewed as a *collectively shared, self-sustaining system of beliefs* regarding a salient way in which the game is repeatedly played. I may identify “a way in which the game is repeatedly played” as the rules of the game. However, by that I do not mean the rules exogenously given or conditioned by the polity, culture, or a meta game, as the rules-of-the-game theorists do. I do regard them as being endogenously created through the strategic interactions of agents and thus self-enforcing, as the equilibrium-of-the-game theorists do. A precise, generic formalization of this view is given in my forthcoming book, *Toward a Comparative Institutional Analysis* (MIT Press).

Institutions as a shared, self-sustaining system of beliefs, may take various forms depending on the game form (the exogenous rules of the game), as well as how the game is played. An example

is given by the Greif-Milgrom-Weingast parable of the merchant guild: “If the ruler cheats merchants, the merchants will boycott his castle town for ever.” This appears to prescribe an action rule for the merchants to follow in the event of a ruler’s cheating. However, such an event may not actually happen, because a ruler’s inference about the merchants’ action-choice rule prevents him from taking that action. Therefore, the rule above actually represents a belief of the ruler. This is an example of regulatory institution in which a credible belief of a certain type of sanction constrains the actual action choice of agents. As suggested, institutions of this type may be analyzed as subgame perfect equilibria of repeated games. On the other hand, since Durkheim sociologists traditionally focused on normative expectations that prescribe certain action choices for all agents in a domain or particular types of agents. However, norms may also be observed by agents because of their beliefs in certain sanctions, such as social ostracization, against default of prescribed obligations. Thus, they may be susceptible to the game-theoretic analysis in similar manner as regulatory institutions. There are also practices and conventions that are constructed by the interactive or mimetic behavior of agents, and simply became “taken for granted.” In these cases, too, however, defection from conventions may be expected to be penalized by utility loss.

As a salient feature of an equilibrium state is objectified as an institution, it may come to be represented in some tangible or explicit forms, such as laws, agreements, social structures or organizations as systems of differentiated roles, routine procedures and so on. However, the point is that a certain representation is an institution only if the agents have beliefs in it. From this perspective, statutory law and regulations *per se* are not institutions, if the agents regard it irrelevant. On the other hand, certain practices, if not formalized, can be institutions as long as the agents believe in them as relevant representations of the internal state of the domain, while they cease to be institutions when the agents’ beliefs in them are shaken. We can treat laws and regulations as setting parameters for defining game forms – that is, specifying exogenous rules of the game – in relevant domains except for the polity domain in which the formation of laws is treated endogenously), and examine what stable outcomes – that is, institutions – they will, or will not, generate *via* its incentive effects on agents. However, it may well be the case that an outcome will be inconsistent with that which a legislature or government initially intended. In such cases, the law or regulation can hardly be regarded as an institution. For example, even if the government prohibits the importation of some goods by a law, but if people believe it effective to bribe customs officers to circumvent the law and make it a prevailing practice, then it seems appropriate to regard the practice rather than the ineffectual law as an institution. As another example, it may not be possible for the legislature to arbitrarily specify enforceable property rights rules for emergent types of intellectual products. Instead, lawyers may take pains to “discover” an emergent, workable praxis among practitioners to formulate an implementable law. This suggests that an emergent practice precedes a codified law if

the latter is to have the power of effective enforcement so that "the game of human society will go on easily and harmoniously, and is very likely to be happy and successful"(Adam Smith).

The summary-equilibrium-representation view of institutions helps clarify their dualistic constraining/enabling nature. The role of institutions is normally understood as constraints on the action choices of the agents (e.g., North). Indeed, an institution, by the very fact of its existence, controls agents' individual action-choice rules by coordinating their beliefs. These beliefs channel their actions in one direction against the many other directions that are theoretically possible (i.e., other equilibria). In this sense, controlling or constraining character is certainly inherent in institutionalization, regardless of whether shared beliefs are regulative, normative or cognitive. However, an institution coordinates agents' beliefs only through a summary representation of an equilibrium state. In a world of incomplete and asymmetric information, an institution "enables" the bounded-rational agents to economize on the information processing needed for decision-making. Here, an analogy with the price mechanism familiar to economists may be useful. In the market mechanism, individuals do not need to know every detail of the internal and external environments of markets in which they make their choices, but only the relative prices (Hayek 1945). Leaving aside the problem of the enforcement of contracts and property rights, if there were a complete set of markets, relative prices could be formally regarded as "sufficient" summarizing the data needed for society to achieve the social optimum in the most efficient way. The dimensionality of relative prices does not exceed the number of goods traded (Koopmans, Hurwicz). Needless to say, in actuality markets are far from complete. Individual agents need alternative means to gain useful information for making their choices. Various institutions other than markets then evolve in response to the failure of complete markets to exist(Arrow 1998). In short, individual agents are not only constrained but also informed by institutions. Just as markets transmit information regarding the economic environment (technologies, tastes and resource endowments) in the summary form of relative prices, so do other institutions in alternative summary forms.

A conceptualization of institutions is of course a matter of the theorist's taste and not a matter of right or wrong. However, in my view there are at least following five reasons why the summary-equilibrium-representation conception of institutions is useful for Comparative Institutional Analysis. The first three reasons are applied to the institution-as-equilibrium-of-the-game view in general, while the last two, more or less specifically, to the "summary equilibrium representation" view.

First, the institution-as-an-equilibrium approach in general can deal with the issues of the origins of an institution and its enforcement endogenously. As we have seen, if one subscribes to the traditional exogenous rules-of-the-game view, then one must immediately face the issues of where and how the rules originated, as well as how they are enforced. Institutional origin may need to be found outside the domain of the economy in which the rules are applied: for example, in the polity

domain outside the economic domain or, theoretically, in the domain of a meta game in which rational agents collectively choose a rule from the set of many possible rules. But how, then, are the rules of the game in the polity domain set? How are all the possible rules known to the players of the meta game and how do they play the meta-game? Where are the rules of the meta game determined? Thus, a problem of infinite regression seems bound to arise. Perhaps the right way to partially resolve this problem is to regard an institution as originating as a stable endogenous product of the game, in either economic, social, or political exchange domains, while leaving the non-technological rules of the game unspecified as much as possible at the outset.

One caveat is due, however. Although we wish to understand institutions as the “endogenous rules of a game” generated in a domain, we cannot build a model to make every possible institution simultaneously endogenous. In other words, what constitutes the “exogenous rules of the game”, i.e., the set of agents, the set of their action choices, the way in which each profile of agents’ action choices is transformed into consequences, may not be completely described by technology, resource endowments, and the preferences of the agents alone. This is the point first addressed by A. Field (1979, 1981). To see the same point from a little bit different perspective, imagine hypothetically that the exogenous rules of some game could be completely specified by technology. Even if it is possible to do so, however, there will be multiple equilibria in a repeated game situation and, as already pointed out, which equilibrium is chosen from them cannot be determined endogenously. We need to consult historical events and rules prevailing in the past, as well as rules prevailing in surrounding domains. It is so, because the particular subsets of actions which agents perceive as the sets of viable options may be constrained by historical precedents, while the way in which the consequences of a certain profile of agents’ actions are determined in one domain may be affected by the institutional environments of the domain (i.e., endogenous rules of the game prevailing in surrounding domains). One can never have an institution-free world from which to start the analysis and completely eliminate appeals to exogenously-given, humanly-devised rule structures. Thus nobody can escape from the problem of infinite regression. However, we may seek to direct the infinite regression toward structures inherited from the historical past rather than the logical construct of the meta-game.

A similar problem of infinite regression can arise with respect to enforcement in the exogenous rules-of-the-game approach. Leaving aside norms and conventions that are self-enforcing (“informal” rules in North’s sense), if the rules of the game (“formal” rules) are to be enforced by an augmented player (enforcer), the question of the enforcer’s motivation needs to be addressed. Who enforces the enforcer, that is, do we need still another enforcer to enforce the rules of action prescribed for the original enforcer? A solution to this problem is again to analyze a game including the enforcer, if any, as a player, and see if the prescribed rules of action for the enforcer can become his/her equilibrium strategic choice and thus self-enforcing, given an equilibrium constellation of

strategic choices by other agents and vice versa. But when we confront the issue of how an enforcer is added to the set of agents, we will not be able to answer the question completely within a game-theory framework, but will need to rely on historical information.

Second, by showing the possibility of multiple equilibria in specific models, the institution-as-an-equilibrium approach is able to shed light on the non-arbitrary, “humanly devised”(North 1990) 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 representation of the technological condition, and not an institution. Usually, a multiplicity of equilibria is regarded as troublesome by game theorists, and they have spent much research effort, without decisive success, in the so-called “refinement” of the equilibrium to enable them to identify only one equilibrium out of the many possible Nash equilibria. However, I 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 the important historical, political, and social factors that selected one equilibrium over the others in particular economies. Thus, equilibrium and historical analyses are mutually complementary and are each indispensable in Comparative Institutional Analysis (Greif). Also, by making institutions susceptible to equilibrium analysis, it can be made clear that institutions are humanly devised, yet can be neither arbitrarily designed nor and discretionary implemented.

Third, the institution-as-an-equilibrium approach provides an appropriate framework for analyzing the interdependencies of institutions operating within the economy. When the government drafts a statutory law for the purpose of introducing a so-far non-existent “institution,” their implementation in particular economic, political and social contexts can often have unintended consequences. For example, even if a government in a transition economy drafts a privatization law aimed at emulating markets for corporate control in the advanced economies, an outcome may be the widespread capture of corporate control by insiders, such as ex-party bureaucrats, directors of ex-state owned enterprises, who amassed de facto control rights before the “transition.” This situation is somewhat analogous to the one in which a medicine which has been tested in a laboratory may have unpredicted side-effects when it is administered to a human being because of the complexity of the living organic system. A major reason for such unintended outcomes is the absence of “fits” between the designed plans and the existing institutional environments which reflect a unique historical trajectory of institutional development. This suggests the possibility that only institutional arrangements that are mutually consistent and/or reinforcing may be viable and sustainable in an economy. It may not be accidental that co-determination in the corporate governance domain and social democratic corporatism in the polity domain co-evolved in Germany, while the main bank

system, life-time employment system, and the close alliance between industrial associations and relevant administrative bureaus co-evolved in Japan, both in contrast to the so-called Anglo-American model.

We can conceptualize possible institutional interdependencies as equilibrium phenomena of linked games (institutionalized linkage) and institutional complementarities (see Aoki, *Toward*, Part II). These intuitively appealing notions can become amenable to rigorous analysis when the equilibrium-oriented notion of institutions is applied. Specifically, we can consider games in different domains of the economy, such as organizational coordination, commodity trade, transactions of services of human and financial assets, 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 equilibrium choices of the same or different agents in another domain. In this way, we can understand the conditional robustness of an over-all institutional arrangement of the economy as well as the multiplicity of such arrangements.

Fourth, the summary-equilibrium-representation view of an institution may suggest a new way to approach the mechanisms of institutional change. As noted already, the ability of an institution to transmit information is not complete. But for the bounded-rational individual agents the compressed information may not only be adequate for making satisfactory strategic choices in a relatively stable environment, but it also provides “backgrounds”(Shearle 1998) for freely and innovatively developing individual skills and competence. Collectively, by providing common compressed information, an institution serves in effect as a mechanism for coordinating the actions of individual agents in the domain *vis-a-vis* their external environment, whether technological, natural, or institutional (in surrounding domains). The subjective adequacy of patterned choices under an institution may become problematic and a “perceptual crisis” is triggered, however, when there is a drastic environmental change, the sustenance of patterned choices becomes severely mismatched with accumulated skills and emergent competence, or more likely a combination of both occurs. The “taken-for-grantedness” of an institution is questioned and the agents are driven to reexamine their own choice-rules based on new information.

A new institution will emerge only when agents’ choices become coordinated in a new manner and their summary representation induces convergent beliefs among them. But this transition may not be just a move from one equilibrium to another for given exogenous rules of the game. Rather it may involve a novelty, characterized by a move from an equilibrium under given sets of action choices of agents to an equilibrium under other (possibly expanded) sets of action choices. In the transition process, various choice-rules involving new actions may be experimented with and put into competition by agents. How can the convergence of beliefs and the coordination of new choices be simultaneously induced in such a situation? The present state of economics has not been able to

show that a dual convergence, both in actual choices and expectations, is possible through reasonable mechanisms of mutual interactions (i.e., actual choices are induced by expectations and expectations are formed by detailed observations of actual choices), particularly when any novelty in choices is involved.

But, it can be through the guidance of a particular, symbolic system presented in the transition process and recognized as ‘prominent’ or ‘salient’ (among the many possible) that agents’ new strategic choices are induced to be coordinated. As agents’ choices equilibrate, the guiding symbolic system becomes consistent with, and reconfirmed by, their experiences. It can then serve as their summary representations and thus it becomes established as an institution. The point is that some symbolic system precedes the evolution of a new equilibrium and then is accepted by all through their experiences. It could be a system of cultural beliefs (Greif 1989, 1994), an entrepreneurial creed, or in some cases even the political program of a political party (as in the Russian or Chinese Revolution). We can describe how boundedly-rational, individual agents form their own subjective models of the game they play, and discuss the mechanism of institutional change as a process of revision, refinement, and inducement of mutual consistency of such models (Aoki, *Toward*, chapter 8)..

Finally, but not least important, the difference in whether the rules of the game constituting institutions are viewed as endogenous to the relevant domain or exogenously set in the polity may have significant implications for the role of public policy. If one subscribes to the view that institutions are constituted of polity-determined rules yet that institutions matter to the performance of an economy, its implications could then be that a badly performing economy should reform itself by designing and implementing better rules, possibly emulating best practices elsewhere. If this was not realized, blame could then be placed on politicians. However, there are two problems with these kinds of arguments. One is that the government itself is an organization of the people who have their own motivations and aspirations. It is an endogenous player of the game in the polity domain and the outcome of any policy advice should be understood as determined by the interactions of the strategic expectations among the players, the government, politicians, and private agents. Secondly, as already noted, a policy may not yield the outcome intended by the government or politicians if it does not have fits with existing institutions, an accumulated stock of competence of agents, and so on. In this book, we will treat laws and regulations as exogenously set parameters for defining game forms (exogenous rules of the game) in domains other than the polity, and examine what the outcomes of the strategic interactions of the agents under them will be. Laws or regulations may induce the evolution of an institution but they themselves are not institutions.

The comparison of the transition from the planned economy to the market economy between Russia and China, as well as the experience of Asian currency crisis in the late 1990s (e.g., Stiglitz,

Qian, etc.), is a pointed challenge to the conventional advice for transplanting the rules of the law practiced in advanced market economies as fast as possible to reform non-market or developing economies. It calls for a careful and systematic study of how the initial institutional conditions, such as the legacies of old institutions and the prevailing informal rules (norms, social ethics, etc.), kinds and level of the existing stock of human competence and so on affect policy impacts on subsequent institutional change, how rule-setting in the polity interacts with the endogenous evolution of the rules of the games in other domains, and so on.

The subtlety of the issue discussed now is that endogenizing the government does not necessarily mean that the game we are studying will become fully determined, leaving no scope for policy advising. As indicated in the preceding premise regarding institutional change, when a certain domain is stricken by the “perception crisis,” individual agents may not have clear strategic expectations, or even if they think they do, their expectations may not necessarily be mutually consistent. Then, there can be latitude for some exogenous symbolic systems to compete for the position of an attractor or a “focal point” (Schelling). They could well be programs or platforms of competing political parties or political advice of the professionals. The point is that, whatever they are, one of crucial factors that will determine their fate is still their “fits” with emergent practices in domains other than the polity.