

Discussion Paper No.288

Masahiko Aoki's Conception of Institutions

Hirokazu Takizawa
Chuo University

November 2017



INSTITUTE OF ECONOMIC RESEARCH
Chuo University
Tokyo, Japan

Masahiko Aoki's Conception of Institutions¹

Hirokazu Takizawa

Faculty of Economics, Chuo University

Email: taki@tamacc.chuo-u.ac.jp

Abstract

The present study examines Masahiko Aoki's continual attempts to conceptualize institutions, with main focus on Aoki (2001) and Aoki (2011). The unique aspect of his approach is identified as his efforts to grasp the dynamic and collective-cognitive nature of institutions. Highly valuing this uniqueness, I identified several difficulties in this approach. In this study, I argue that the profound nature of institutions is beyond the formal description of game theory, which Aoki maintained in his lifetime. I also that the Hegelian perspective submitted by Herrmann-Pillath and Boldyrev (2014) suggests a possible future direction of the research on institutions in economics, albeit not necessarily one amenable to mathematical formulation.

JEL Classification code: B31, B41, B52, C70, D02, D80, P51

Keywords: institutions, conception of institutions, Aoki

¹ The author is grateful to Carsten Herrmann-Pillath for invaluable discussions on Professor Aoki's approach to institutions. Thanks also go to the attendees of the session commemorating Masahiko Aoki at the Annual Meeting of Japan Association for Evolutionary Economics, March 26-27, 2016, Tokyo. This study was supported by JSPS Kakenhi Grant Number 16K03557 and the MEXT-Supported Program for the Strategic Research Foundation at Private Universities 2013-2017.

1. Renewed Interest in Institutions in Economics

Since its very inception in the 18th century, economics has paid considerable attention to the importance of institutions, as best exemplified by the economic policy proposal of Smith (1759/1976; 1776/1991). However, this interest diminished as the neoclassical theory established itself as mainstream in the 20th century, whereby most economists focused exclusively on the workings of the market institution/mechanism. This approach succeeded because the marginal gain from this type of study was extremely high.

The late 20th century saw this trend being reversed for several different reasons. First, as the economic advantage of capitalism over socialism became obvious in the 1980s, economists' attention shifted toward the difference in institutional arrangements between major capitalist economies, which seemed to exert great impact on their respective countries' economic performances. Second, the experience of transitional economies after the collapse of some socialist regimes showed that the market does not stand on its own, but its performance is actually supported by various non-market institutions, such as the institution of property rights, commercial law, and so on. However, the most important factor came from within economics itself. The early 1970s experienced a theoretical revolution in economics, which directed our attention to how institutions are set up to alleviate problems arising from "asymmetric information" (Akerlof 1970). In the 1980s, almost all works in this field came to be rewritten by using game theory.

Of course, important works in institutional economics as well as in game theory already existed well before the 1980s. However, the fruitful marriage between the two meant that we became able to talk about the workings of diverse institutions (not only of the market) by rigorous mathematical standards that most economists were satisfied with². This achievement has been welcomed by traditional economists, since it helped us understand how various non-market institutions support the market mechanism (McMillan 2002), which is the central theme of classical economics.

In retrospect, this extension of the economists' capacity to analyze institutions led to another extension of their interests. Institutional phenomena are ubiquitous in our everyday life, deeply rooted in our social/behavioral dispositions. Combined with the rise of behavioral economics, it was then that some economists began to pay attention to human

² The reason that game theory is helpful for understanding institutional phenomena is that the formal definition of games has a strong expressive power—it allows us to model fine details of asymmetric information, gives predictions of the observed regularity of behavior, and understand the role of belief in sustaining the equilibrium behavior.

social behavior constituting institutional phenomena. This shift obviously called for a transdisciplinary approach, because there are many disciplines outside economics concerning human behavior/sociality related to institutions. Therefore, some, if not all, economists came to show their interest in the research results in those areas.

It is against this backdrop that several economists began to pose a question about the general nature of institutions from their own perspectives³. The above described nature of the renewed interest in institutions, however, exerted enormous impact on those attempts: under the overwhelming influence of game theory, almost all conceptions of institutions so far proposed in economics borrowed their defining terms from game theory.

Thus, one strand of these attempts regards institutions as “the rules of the game.” This approach, henceforth called “institutions-as-rules” view, is best represented by North (1990), which begins with a famous passage: “Institutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction” (pp.3-4). He also identified the importance of “informal rules,” such as codes of conduct, norms of behavior, and conventions. Institutions are thus supposed to channel and incentivize human behavior, while constantly changing the shape of society and economy until today. The other strand regards institutions as the equilibria of the game (we will henceforth call it “institutions-as-equilibria” view). This approach is already found in an early contribution by Schotter (1981), which defines institutions as “regularity in behavior,” but later joined by Greif (1994).

The institutions-as-rules approach has its strength in explaining the selection process of institutions. Despite North’s important distinction between formal and informal rules, attention was mainly concentrated on formal rules created by a polity, since it is difficult for policy makers to affect informal rules. The process was supposed to be ultimately driven, rather exogenously, by the institutional entrepreneurs’ motivation to change property rules to their advantage. Whereas North himself emphasizes the importance of enforcement, this approach does not explicitly explain how institutions are stabilized in a society, which may require understanding of how people’s behavioral regularity emerges.

On the other hand, the institution-as-equilibria view has the advantage of being able to explain how institutions often have long-standing stability. This is mainly because this view

³ Hodgson (2015) correctly points out that the task of defining institutions is conceptually different from the task of understanding or analyzing them. Searle (2015) warns us that we must be careful when we ask questions of the form “what is...?” With these precautions in mind, I take up this subject matter, because I believe that how we conceptualize institutions is important, as it directs our research, by defining what kind of phenomena we regard as important and want to understand.

identifies an institution as the Nash equilibrium of a relevant game, which has the so-called “self-enforcing property,” whereby any unilateral deviation does not benefit a player. Thus, enforcement is endogenous in this approach. This view also enables us to understand why there is diversity of institutions in similar domains across economies, by resorting to the existence of multiple equilibria of the relevant games. However, this approach also has its own weakness. While it is very amenable for explaining how institutions are maintained/stabilized, it is hard to explain how institutional changes occur, and we sometimes observe institutional creativity/novelty in the changing process. The understanding of this process is indispensable if economists are to take the explanation of historical changes seriously.

It is in this context that Masahiko Aoki started to elaborate his own conception of institutions, which is sometimes simply understood as belonging to the institution-as-equilibria view. However, that would prove to be a misunderstanding if one carefully reads his books/papers. What motivated him to examine this difficult question was his desire to overcome the weakness of the equilibria view. As we will see later, he wished to understand the process of institutional changes/novelty in a way consistent with the equilibrium view of an institution. What really interests us is what he invented to address this conundrum.

Some qualifications are in order at this juncture. While the present study focuses on Aoki’s theoretical conception of institutions, it will not deal with his ideas on policy issues. This restriction is actually unfortunate, since he never lost interest in political issues, and his notion of institutions is intertwined with his ideas on policy issues. However, discussing the relationship between his theoretical conception and his view on policy issues would be beyond my ability and thus outside the scope of this study. This may be partly justified, since his conception of institutions was increasingly directed toward understanding the “institution in deep structure” in his later writings.

The organization of the paper is as follows: Section 2 closely examines Masahiko Aoki’s continual attempts at conceptualizing institutions, with focus on Aoki (2001) and Aoki (2011). Section 3 looks at a debate that took place in the *Journal of Institutional Economics*. Seemingly the debate revolves around a topic paper submitted by Hindriks and Guala (2015). However, their view, called the “rules-in-equilibrium” view, is essentially Aokian in its spirit. Therefore, the debate incidentally illuminates the difficult problems faced by Aoki’s conception of institutions. Section 4 first discusses those difficulties in some detail, and suggests a possible solution, by referring to Herrmann-Pillath and Boldyrev’s (2014) Hegelian approach to institutions. Finally, Section 5 concludes the paper.

2. Masahiko Aoki's Conception of Institutions

Let us now closely examine Aoki's conception of institutions. He began to seriously engage in this research question when he was writing *Toward a Comparative Institutional Analysis* published in 2001 and his endeavor continued until just before his death in 2015. In the beginning of the 2010s, he was deeply involved in setting up a truly independent policy institute in Japan (RIETI: Research Institute of Economy, Trade and Industry), where he personally tried to understand the seemingly endless impasse of the Japanese society/economy, usually called the "lost decades," as a process of transformation of institutional legacy that supported the post-war Japanese regime. In 2007, after leaving RIETI, he set up another research institute at TKFD (Tokyo Foundation), called VCASI (Virtual Center for Advanced Studies in Institutions), to which he invited cutting-edge scholars from various fields related to human behavior and society, including cognitive scientists, brain scientists, philosophers of the social sciences, and sociologists, among others, and had intense discussions with them. To the best of my knowledge, this theoretical effort to conceptualize institutions was crystalized in Aoki (2011), where he gathered all of what he learned through the research activities in VCASI⁴.

Therefore, I will choose to focus on his ideas expressed in Aoki (2001) and Aoki (2011). His ideas evolved somewhat between these two writings, although his basic ideas remained the same.

2.1 The Conception of Institutions in Aoki (2001)

As is well known, Aoki (2001) is a serious attempt to launch and give foundations to a new field in economics: Comparative Institutional Analysis. In the beginning, Aoki sets two goals for this subfield: one for the synchronic problem and the other for the diachronic problem. The goal of the synchronic approach is "to understand the complexity and diversity of overall institutional arrangements across the economies as an instance of multiple equilibria of some kind," whereas that of the diachronic approach is "to understand the mechanism of institutional evolution/change in a framework consistent with an equilibrium view of institutions, but allowing for the possibility of the emergence of novelty" (pp.2-3).

Then, he reviews and summarizes several conceptions of institutions hitherto proposed, the institutions-as-rules view and institutions-as-equilibria view, and embarks on submitting

⁴ Whereas Aoki (2010) tries to understand and exposit the evolving diversity of corporations and is also relevant, the idea presented there is basically the same as in Aoki (2011).

his own “equilibrium-summary-representation” view of institutions. The first definition he gave deserves quoting at length:

[W]e will tentatively characterize the institution as a self-sustaining system of shared beliefs about a salient way in which the game is repeatedly played. We can identify ‘a way by which the game is repeatedly played’ as the rules of games. However, by that we 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. We regard these rules as being endogenously created through the strategic interactions of agents, held in the mind of agents, and thus self-sustaining—as the equilibrium-of-the game theorists do. In order for beliefs to be shared by the agents in a self-sustaining manner and regarded by them as relevant to the consequences of their choices, they must have substantive bases. The content of the shared belief is summary representation (compressed information) of an equilibrium of the game (out of the many that are theoretically possible). That is to say, a salient feature of an equilibrium may be tacitly recognized by the agents, or have corresponding symbolic representations outside the mind of agents and coordinate their beliefs (p.10).

At first reading, this definition sounds very similar to the “institutions-as-equilibria” view, toward which he was obviously sympathetic. That being the case, what makes his definition distinct from the ordinary institutions-as-equilibria view? First, note that this definition does not simply focus on the “regularity of behavior” or equilibrium, as most game theorists might do, but it identifies an institution as a system of “shared beliefs” held in the mind of agents. Second, he tries to attract our attention to the role of “representation” outside the mind of agents in coordinating their beliefs. He even says that “the dashed-line box represents an institution” in Figure 1. This means that he thought extensively about the cognitive aspects of institutions, especially agents’ capability for representation, in order to allow for institutional emergence and novelties, which are the second goal of his book.

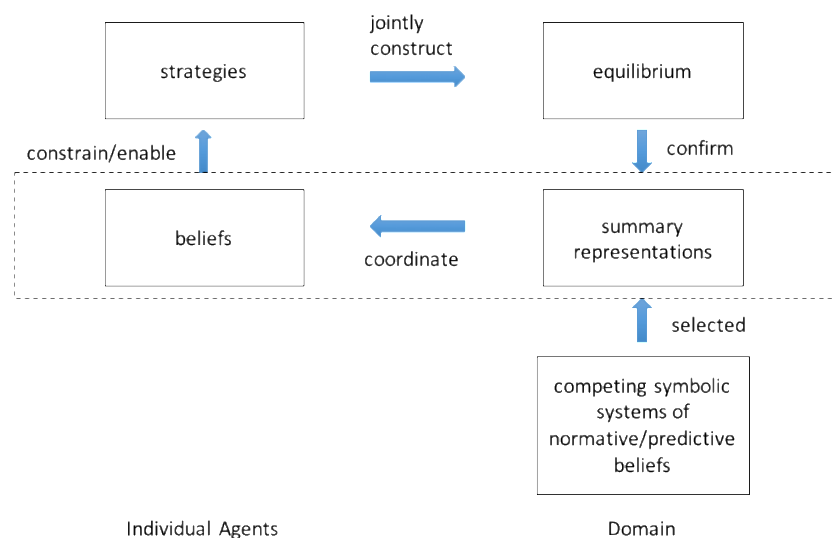


Figure 1 An institution as shared beliefs formed as a summary representation of an equilibrium. An institution is represented by the broken-line box (Aoki 2001, p.12)

Somewhat surprisingly, starting in 2001, he had already alluded to the importance of something outside agents for coordinating their beliefs, which he calls “symbolic representations” here. Therefore, what we see in his definition is a reconciliation of “internalism” and “externalism”⁵.

An ordinary game-theoretic analysis of an institution would define a well-defined game describing the relevant situation, and solve for equilibria to deduce implications. Each equilibrium of the game specifies a certain strategy profile and a set of agents’ beliefs thereof. At an equilibrium, the beliefs of the agents (lower-left box) support their strategy choices (upper-left box), which then constitute an equilibrium (upper-right box). Subsequently, the equilibrium will in turn endogenously determine the beliefs of agents, which brings us back to the lower-left box. Then, the game-theoretic analysis is completed with the three boxes in Figure 1. Compared with this framework, where Aoki’s interests reside should be clear: He was interested in both the coordinating/self-sustaining and changing/creative nature of institutions.

This description necessarily involves some more deviation from the ordinary games that game theorists deal with. The rule Aoki mentions in the above quote is not what a game

⁵ “Internalism” here refers to a stance in the philosophy of mind that regards actions as external phenomena to be explained by resorting to internal determinants, such as preferences and beliefs, whereas “externalism” grasps the workings of a mind more broadly, and thus emphasizes the interaction between brain, body, and external environments.

theorist would have in mind. For a game theorist, the rules of a game consist of a set of players, strategy sets for each player, and players' payoff functions, which map the set of possible strategy combinations into real numbers. The rule Aoki mentioned above is a rule of conduct closely related to an equilibrium, which makes his position close to the equilibrium view on the one hand, but it also appears as external constraints to individual agents, namely, as rules in the institutions-as-rules view. However, this stance has to assume that agents are playing a gigantic game where their strategy set consists of "all technologically feasible actions" (p.186) in terms of formality. Of course, the agents here are not supposed to be able to know every detailed characteristic of others' actions. Therefore, a summary representation is needed so that they only perceive the salient features of the behavioral rules⁶.

2.2 The Conception of Institutions in Aoki (2011)

The last decades of the 20th century were the period of a resurgence of interdisciplinary discussions between economists, cognitive scientists, and computer scientists, which still continue until today. For example, Denzau and North (1994) heavily relied on the insight of externalist arguments developed by cognitive scientists to enrich their theory of institutions. In his 2001 book, Aoki frequently mentioned Denzau and North (1994), but the references were not substantive, but mainly used for undergirding his own concept of "subjective game model" and "shared belief." However, given the interest he maintained for institutions since 2001, it was natural for him to look for arguments outside economics, which are instrumental in explaining how the agents' beliefs are coordinated. In his research activities at VCASI, he paid much attention to the literature on externalist arguments found in philosophy and cognitive sciences, such as Heath (2008) and Clark (1997)⁷. This process was crystalized in Aoki (2011), on which I will now focus.

As the title of the 2011 paper, "institutions as cognitive media between strategic interactions and individual beliefs," suggests, Aoki identified more clearly the quintessential role of institutions in mediating agents' strategic interaction and their individual beliefs in societal games. This obviously shows the continuity of his basic position, but more emphasis was put on the *external* character of institutions. To quote from the abstract, that study tries

⁶ This is the aspect embraced by Hendriks and Guala's (2015)s "rules-in-equilibrium" account of institutions, as we will see later in Section 3. Chapter 7 of Aoki (2001) is devoted to the elaboration of this view.

⁷ Aoki recommended both of these books for Japanese translation and they were translated and published in Japanese.

to elucidate “the role of institutions in *substantive forms as societal artifacts* that cognitively mediate agents’ strategic interactions and their individual beliefs in societal games” (emphasis added by the present author). He named this new idea “institution-as-cognitive-media-view.”

What are the societal artifacts that can assume the role of cognitive media? He found the answer to this question in the classical, but often overlooked reasoning of Lewis (1969). Lewis (1969) identifies the essential problem of conventions in the coordination games with at least two proper coordination equilibria⁸, and argues that a system of concordant higher-order expectations should be established for solving the coordination problems. The subsequent research in game theory showed that the argument does not have to be confined to coordination games, but applies to equilibria in games in general (Aumann 1976; Aumann and Brandenburger 1995).

The original argument proceeds as follows: Consider that the following state of affairs, denoted as A , has occurred; you and I were discussing an important business problem and you had to leave during the discussion. You promised to come back the following day. This event, A , is supposed to satisfy the following conditions:

- (1) You and I have reason to believe that A holds,
- (2) A indicates to both of us that you and I have reason to believe that A holds, and
- (3) A indicates to both of us that you will return.

Here, the meaning of “indication” plays an important role. A “indicates” to someone, x , that B holds, if and only if, x has reason to believe that A holds; then, x has reason to believe that B holds. Assuming, further, the common inductive standard and common background information between the relevant agents, the self-referential nature of Condition (2) ensures that arbitrarily high-order beliefs are established between you and me. For the pivotal role that A plays in creating a common knowledge, Lewis called it a basis of common knowledge.”

Aoki found this to be the theoretical entity that ensures the coordination of the agents’ beliefs, and identified it as “social cognitive artifacts” or “public representation.” Suppose that some state of play (strategy profile) a occurs in the set of all possible states of play A . Each agent has his/her own information function $P_i(a)$, meaning that when a occurs, he/she only knows that some element in $P_i(a)$ occurred. Let $m(a)$ be the element of the

⁸ Lewis’s definition of proper equilibrium is the same as what we usually call “strict equilibrium”; he calls a strategy combination “a *proper* equilibrium if each agent likes it better than any other combination he could have reached, given the others’ choices” (Lewis 1969, p.22).

meet⁹ of the agents' information partitions containing a . Now suppose that some public proposition P^* has the following properties in parallel with Lewis's argument:

- (1) Every agent in the society has reason to believe at a that P^* holds (mutual knowledge);
- (2) Every agent infers from P^* that possible states of play are in $m(a)$;
- (3) P^* indicates to every agent that every other agent also has reason to believe that P^* holds and derive the same inference from it (symmetric reasoning).

Then, Aoki claims, we can derive that P^* indicates to every agent that all agents believe that all agents believe ... that the possible states of play are in $m(a)$. In this way, $m(a)$ becomes a common knowledge among agents if and only if P^* holds. If P^* summarily represents the salient feature of recursive states of play in society, this enables agents to coordinate their beliefs (See Figure 2). To quote Aoki himself:

A public proposition P^* , mediating the stable physical states of play (strategic interactions) and individual beliefs in recursive ways, may be referred to as a *substantive form of an institution* ... It summarily represents the aspect of recursive states of play of the societal rules as captured by $m(a)$ and thereby induces the partial convergence of individual behavioral beliefs toward $m(a)$, which in turn reproduces the states of strategic play to fall in $m(a)$ over time and so on." ... As a social cognitive medium, it is bound to take a linguistic form. For the sake of convenience, I tentatively refer to this view of institution as the *institution-as-cognitive-media-view*. (p.25)

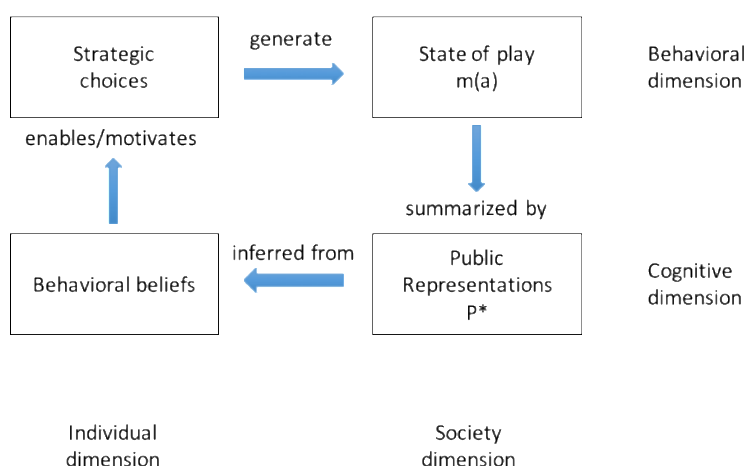


Figure 2 The mediating role of institutions in substantive form (Aoki 2011, p.25)

⁹ A meet is the finest common coarsening of agents' information partitions.

However, how can this process be realized in reality? Aoki (2011) answers this question in Section 4, where he mentions the crucial role of culture “as encapsulating the experiences of past generations of any particular group” (Denzau and North 1994, p.15, quoted in Aoki 2011, p. 26). The discussion in this part is extremely difficult to follow, but he seems to identify the two-fold role of culture in this process: On the one hand, culture works as a common prior over the set of all possible types of agents involved, from which each agent assesses the posterior probability regarding the types of other agents by a conditional distribution, given his or her own type. Culture also provides a basis for similar inductive standards among agents assumed in Condition (3) above.

However, there seem to be several difficult issues here. First, Aoki seems to suppose that emergent states of play may even precede a rule of the game as usually defined in game theory. Indeed, he focuses on behavioral rules (states of play) without considering a rule of the game. He states that they are recursively observed “even if they [agents] do not know more in common (including the objective structure of game)” (p.23) and calls those observed behavioral rules *institutions in deep structure*. Understandably, this step aims at reconciling the dichotomy between the institutions-as-rules view and institutions-as-equilibria view by endogenizing the rules of the societal games. However, to do this in a rigorous analytical framework requires him to define A as the set of all possible states of play perfectly known only to an outside observer, an unnatural practice also observed in his attempt in Aoki (2001)¹⁰.

Second, he assumes that bounded-rational agents can cognize states of play only in summary forms as $P_i(a)$, resorting to the argument put forward by Hayek in his *Sensory Order* (1952)¹¹. This may partly be the reason that he thought P^* had representation in linguistic forms as in the above quote. However, if the cognitive aspects are so important for the emergence of institutions, they should not be presupposed *ex ante*. That is the reason that he finally turned his attention to the crucial role of culture. However, culture seems to

¹⁰ Aoki’s stance towards rigorous formulation is somewhat ambivalent. On the one hand, he writes “I rely mainly on game-theoretic language and frame of thought with liberal re-interpretations” (Aoki 2011, p.21). On the other hand, mentions his conception of institutions “using rigorous set-theoretic notions” (Aoki 2015, p.486).

¹¹ To quote, “[o]ne of the basic ideas in the book was that the essence of psychological process and knowledge as its outcome lies in the ‘classifications’ of perceived physical events in one way or another through the network of neurons. The formalization of the idea of classification is precisely at the basis of the notion of common knowledge in the contemporary theory of knowledge” (Aoki 2011, p.24).

be an important instance of institutions. Resorting to culture only means that institutions are historically shaped and thus shared in society, resulting in circularity of argument.

3. A Debate on the Conception of Institutions

In 2015, the *Journal of Institutional Economics* put together special features concerning the concept of institutions. The feature issue consisted of a theme-setting study by Hendriks and Guala (2015) as well as commentaries on the paper, contributed by economists/game theorists, Robert Sugden, Ken Binmore, Geoffrey Hodgson, Vernon Smith, Masahiko Aoki, and a philosopher, John Searle. In the topic paper, Hendriks and Guala claim to have developed a unified approach to institutions—what they call the “rules-in-equilibrium” account, that reconciles not only the hitherto proposed approaches to institutions (institutions-as-rules and institutions-as-equilibria, both of which we have already discussed), but also the constitutive rules approach as developed by John Searle.

The constitutive rules view of institutions had been put forward by John Searle in his books and articles (1995; 1998; 2001; 2005), but by then had attracted little attention from economics. Searle asserted that an institution is “any collectively accepted system of rules (procedures, practices) that enable us to create institutional facts.” These rules typically take the form, “X counts as Y in context C,” where the X term is an object, a person, or state of affairs, the Y term a status function assigned to X, and C a context. In order to further understand Searle’s claim, we have to know more about the characterization he gives. The rule here has to be “collectively accepted” in the sense that it is an instance of collective intentionality^{1 2}. “Assignment of status function” means that the object or person cannot perform the function just in virtue of its physical structure, but can perform the function in virtue of the collective acceptance of the object as having a certain status. This differentiates institutional facts from brute facts. When the practice of the form “X counts as Y in C” is regularized as a rule, it will be *constitutive* of institutional structures. Thus, the rules of this kind are called constitutive rules in contrast with regulative rules of the form “do X,” or “if Y, do X.” Constitutive rules enable humans to create new sorts of power relationships—deontic power—which Searle regards an essential characteristic of institutions.

^{1 2} Intentionality is a technical term in philosophy that denotes “aboutness” or “directedness” of the mind and/or linguistic expressions. For example, we cannot believe something without any object to be believed. Those states of mind are called “intentional states.” “Collective intentionality” means that we have common intentionality.

We now return to Hindriks and Guala's account. In the first half of the paper, they consider the private property game (hawk-dove game) to find that institutions must somehow be *correlated* equilibria of *coordination* games with multiple equilibria. From the view point of players, each equilibrium specifies a "rule" to follow. Thus, rules are in equilibrium, in a sense. However, they are not fully satisfied with this definition, because they want to further consider the institutional creativity that only humans can exhibit. They then introduce *representation* as the third condition of institutions, since without representation, humans cannot invent a new institution. Finally, they define institutions as "rules in equilibrium, where the rules are summarized by the agents, using some kind of symbolic representation." In their argument, Aoki (2007; 2011) is cited as emphasizing the importance of public representation or social cognitive artifacts.

In the latter half of their paper, Hindriks and Guala also attempt to integrate Searle's constitutive-rule account of institutions into their framework. Since they assume that actions or strategies appearing in game theory take the form of regulative rules, it suffices to reduce Searle's constitutive rules to regulative rules. To do this, they break down the Y term in Searle's definition of a rule to "base rule" and "status rule." The base rule is of the form "X is Y," and the "status rule" is of the form "X is collectively accepted as Y." They then argue that collective acceptance can be interpreted as an equilibrium behavior in game theory.

In sum, Hindriks and Guala (2015) embrace Aoki's conception of institutions in their rules-in-equilibria approach. Their approach is also similar in spirit to Aoki's in that they introduced representation as an essential part of institutions to take account of institutional novelty. This is the reason that Aoki (2015) sends a friendly comment on their study. However, there is an important difference between the two approaches. Aoki chooses to delve into the collective-cognitive process through which an equilibrium is achieved in a societal game, identifying the essence of institutional dynamics and creativity therein. In contrast, Hindriks and Guala think they can dispense with this issue^{1 3}.

4. Limits of Economic Approaches and a Possible Solution

So far we have reviewed the conceptions of institutions put forward by Aoki as well as others, as long as they are related to his argument. We have also reflected upon the debate that took place in the *Journal of Institutional Economics* in 2015. Consequently, what can we learn

^{1 3} In footnote 14 of their paper, they write "the reducibility of collective to individual intentions is a thorny issue in the philosophy of action." However, they proceed by saying "it is perfectly possible to discuss the relation between the rules-in-equilibrium and constitutive rules approach while remaining neutral on this matter" (Hindriks and Guala 2015, p.471).

from all of them? Where do we stand in the attempt to conceptualize institutions? I find two difficult problems facing economists' current approach to institutions. One is concerned with how to deal with normativity and the other with how to deal with sub-consciousness.

4.1 The problem of normativity

The fundamental modeling strategy of standard economics has been to assume that agents are instrumentally rational, in the sense that they make decisions in a self-interested manner, which brings the best outcome. Agents are supposed to do this by combining two intentional states: their preference and belief. This assumption is so deeply rooted in the research practices of economics that we cannot imagine how to do without it. The game-theoretic approach also follows this principle, since a Nash equilibrium is defined as a strategy profile where each player chooses his or her best response to the strategies of other players. It seems to me that this makes it difficult for economics to explain the normative dimension endogenously.

In the present context, the institutions-as-rules view avoids this issue by supposing that any normative power is somehow reflected in the set of actions for agents^{1 4}, whereas the institutions-as-equilibrium view would hold that it is just a derivative of the equilibrium play of the game^{1 5}. However, is it really justifiable to leave the normative aspects out of our explanation/understanding of institutional behavior?

Some may argue against this by saying that the agents are not simply assumed to be self-interested, because their preferences are defined over strategy profiles rather than simple physical outcomes so that game-theoretic modeling can reflect the phenomenology of the agents' normativity. The results of experimental game theory, however, have shown that subjects in a laboratory usually consider the payoffs of other players. This means that human subjects cannot help considering others' payoffs even when they correctly represent the game situation. In response to this, Heath (2008) persuasively argues that confining ourselves only to two intentional states, preference and belief, is a fully justified practice, when one examines the history of philosophy, especially when we consider agents' behavior in social contexts.

^{1 4} Recall North's definition of institutions: "In the jargon of the economist, institutions define and limit the set of choices of individuals" (North 1991, p.4).

^{1 5} Sugden (2015) does not appreciate Hindriks and Guala's or Aoki's efforts to incorporate representation in the conception of institutions, contending that the regularities of behavior and affective response explained by evolutionary game theory are sufficient as a scientific explanation of conventions.

Since the latter half of the 20th century, we have accumulated plenty of evidence—mainly in primate ethology, developmental psychology, and evolutionary psychology—which indicates that human rationality is the result of some normativity (see Tomasello 2009 for an example). Humans could not even have acquired their linguistic capacity without the imitative and norm-conformative propensity (Heath 2008). This insight was also put forward by Hayek (1991)¹⁶. This is the basic background against which the “social contract” approach to institutions, which regards institutions as contracts concluded by rational agents, is increasingly losing ground among social scientists (see, e.g., Fukuyama 2011). John Searle’s harsh criticism against the game-theoretic approach is also based on this intuition, although for now I will not deal with the issue concerning the distinction between “normativity” and “deonticity.”

That said, however, the above argument does not simply imply that economics cannot properly deal with all institutional phenomena. Economists’ renewed interest in institutions in the latter half of the 20th century mainly originated in understanding the stability or enforceability of institutions, and then shifted toward their emergence and changes. There are many interesting institutional phenomena that economists can explain by making use of game-theoretic reasoning, such as institutional complementarity, linked games, historical path dependence, etc. We would not be able to design any mechanism without considering its enforceability. The point is, however, that any usual game-theoretic modeling presupposes a full specification of the rules of the game, the set of players, and the set of strategies and payoff functions for each player, which is only possible with the prior existence of our common understanding of some institution. This is partly the reason that Aoki (2011) and Gintis (2009) mention culture as a social device that provides a common prior among agents. However, this is the point where the “spade turns” immediately, when economists try to dig more deeply.

4.2 The problem of sub-consciousness

A second difficulty that we are faced with concerns the role of sub-consciousness in human creativity. Since Aumann’s (1976) seminal contribution, it has been a common practice for economists to formulate the agents’ knowledge within a set-theoretic framework, where an agent’s knowledge is expressed by a partition of the set of all states of the world. We have seen an instance of this formulation in Aoki (2001; 2011). Since this model is known to be equivalent to a version of modal logic with knowledge operators, it has some affinity with the

¹⁶ To quote, “Just as instinct is older than custom and tradition, so then are the latter older than reason: custom and tradition stand between instinct and reason” (Hayek 1991, p.23)

idea that institutional phenomena are somehow related to representation.

However, the notion of representation seems somewhat ambiguous. It can mean some internal states of the mind as well as external entities denoting something else. As we have seen, Aoki's "public representation" plays an important role in bridging across these two aspects: it is some (external) state of affairs observable among agents, which however leads to common-knowledge beliefs internal to them. However, the representation thus created does not seem to create anything new, as long as it is just a representation of the "recursive state of the play." This means that, while Aoki succeeded in explaining how an institution is sustained with the idea of representation, he has not yet achieved the second goal in his book (Aoki 2001): to understand how a novel institution is created.

From this viewpoint, it seems strange that Aoki's account does not borrow the useful insights on the mechanism of mind uncovered by the recent developments in psychology—not only that the interaction with the outside world is an indispensable aspect of the mind, but also that sub-consciousness plays an important role in the information processing of the mind (Clark 1997).

Admittedly it is hard to rigorously define what consciousness or sub-consciousness is. However, we now know that we are processing an enormous amount of "tacit knowledge" every moment in our lives (Polanyi 1966). Shimojo (2008) presents a new model of knowledge, whereby both consciousness and sub-consciousness work based on the recent findings in psychology. This is a model developed for explaining the cognitive process of discovery/invention that takes the opacity of knowledge seriously. Shimojo turns our attention to the fact that we can immediately notice that some idea, invention, or scientific discovery, submitted by a genius, is genuinely ingenious, whereas we have not realized it ourselves until that point in time. The classical model of knowledge, say that of Plato, somehow presupposes the set of true sentences with full representation and supposes that, although we may have not noticed, it was just "recalled." Shimojo argues that this model of knowledge is misleading and proposes instead that human knowledge has a two-tier structure, consisting of conscious representation and sub-conscious representation. In this account, in a society, we share sub-conscious representation by sharing common environments. This can explain the fact that we find true ingenuity in the idea submitted by a genius. This story is interesting because it points to the possibility that human creativity lies in our capacity to deal with the dual structure of representation: conscious representation and sub-conscious representation.

This model may not provide us with the neatness of the standard mathematical model of knowledge but seems to me as a strong candidate idea for explaining the institutional creativity humans exhibit.

4.3 A possible solution

At this juncture, Herrmann-Pillath and Boldyrev's (2014) bold attempt to introduce a Hegelian perspective to economics deserves mentioning, since it contains an elaboration of Aoki's model of institutions and possibly represents a solution to the above-mentioned problems.

A remarkable aspect of the recent development of economics is that it has now come to accommodate a "naturalistic" explanation of human behavior/institutions, as best exemplified by the frequent use of experimental methods in behavioral economics and neuroeconomics. This is of course in contrast to conventional methodology of standard economics, which presupposes the rationality of economic agents. From this viewpoint, institutional economics, another emergent subfield of economics, may be regarded as lying in between, since it has to be faced by the deepest aspects of human nature as well as the rationality observed in public discourses. Then, a problem arises, concerning how to totally understand the varying spectrum of human activities and institutions. According to Herrmann-Pillath and Boldyrev, this is the very problem that Hegel tackled in his life-long philosophical endeavor.

The latter authors summarize the essence of Hegelian ideas in three principles: continuity, performativity, and recognition. The continuity thesis asserts that the natural and the spiritual world are intertwined with each other and thus cannot be neatly separated, implying that no naturalistic reduction is possible. Viewed from this perspective, institutional reality is a "second nature." The performativity thesis contends that institutional reality is a dynamic manifestation of the subjective spirit. Thus, human beings can create institutional facts as distinct from the brute facts, the point stressed by John Searle. The recognition thesis asserts that institutions cannot be established without mutual recognition of each agent by one another and of the institution (norm) as such.

Herrmann-Pillath and Boldyrev find an advantage of Aoki's conception of institutions in its emphasis on the totality of the dynamic process covering all the four boxes in Figures 1 and 2. Thus, properly reinterpreted, it can serve as a model encompassing the three Hegelian principles above. To do this, they modify Aoki's original model in several ways (Figure 3). They use the term "sign systems" instead of "public representation" in the lower-right box, and suppose that the right boxes and the arrow between them contain a complex recognition process in the form of "distributed cognition." It is also a process of "meaning making," reflecting Hegel's "expressivist" notion of actions¹⁷. Thus, they do not

¹⁷ In the expressivist view, actions are primary facts and therefore cannot be ontologically separated from their causes/intentions. Rather, actors are engaged in reflective activities, always

have to resort to the game-theoretic notion of common knowledge, or “shared belief” that Aoki often referred to and is rationally deduced from the existence of public proposition as in Figure 2. Therefore, Herrmann-Pillath and Boldyrev do not have to use the term “belief,” an intentional state, in the lower-left box. The arrow from the right box to left one in the lower part, as well as that from the lower-left box to the upper-left box, reflect natural causation, where some neuronal mechanism is working sub-consciously. Game theory also plays a role in aggregating individual actions into recursive states of play in a society, in the upper part.

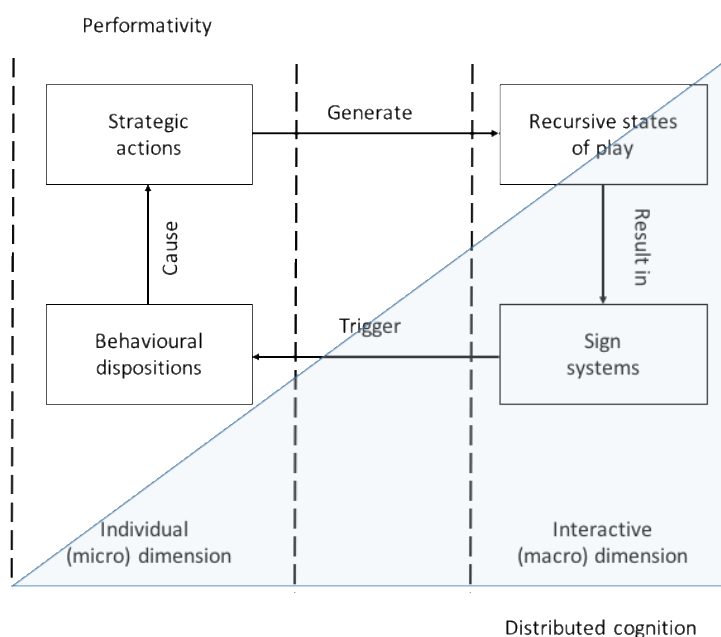


Figure 3 The revised Aoki model (Herrmann-Pillath and Boldyrev 2014, p.71)

They also assert that identity and preference of agents are generated through this dynamic process. Thus, the deontic power identified by John Searle as the “key feature of human institutions” (2015, p.508) is dealt with the concept of endogenous preference. Thus, as long as we do not stick to the primordial/autonomous role of deonticity as distinct from general normativity, this approach can accommodate the normative aspect as mentioned as above. This view is also supported by Aoki (2015). Criticizing Searlian emphasis of deonticity, he writes “I would rather subscribe to the Hegelian-like view of social values as arising from mutual recognitions among agents” (Aoki 2015, p. 487), calling this approach “neo-Hegelian.”

discovering their own intentions related to their action.

I wonder if economists like this framework, since it can hardly be captured by a neat mathematical model, but I find this approach to be the most promising one as an elaboration of Aoki's conception of institutions^{1 8}.

5. Conclusion

The recent decades have seen economics embracing such unconventional research fields as experimental economics, behavioral economics, and neuroeconomics, as exemplified by the awarding of the 2002 Nobel prize in economics to Vernon Smith (experimental economics) and Daniel Kahneman (behavioral economics). This process has not only expanded the scope of economics, but has exerted an enormous impact on the methodology of economics as well as on how economists think about the relationship between their theory and the real world.

First, this process introduced “naturalistic” methodology to economics; by now we have economic researches that investigate how economic agents think and behave in the *real* world, instead of just assuming they follow certain axiomatic rules, such as utility maximization. Secondly, related to the first point, economists (at least some of them) have come to pay more attention to interesting economic phenomena in the real world. The economic theory of institutions has also been under this influence.

Developed in the middle of this process, Aoki's theory of institutions was a bold attempt to integrate social/behavioral sciences, whereby institutions serve as a hinge that connects extensive subfields. It should be noted however that he himself tried to remain within the game-theoretic framework in his style of research.

His attempt was very hard, because, on the one hand, there still are strong propensities on the side of economists toward mathematical modeling, and on the other hand, we have by now accumulated a host of interesting institutional phenomena that cannot be grasped by the current game-theoretic framework. Probably, this situation will continue for some time. Hence, is it totally meaningless for economists to try to conceptualize the essence of institutions? I think the answer is “no.”

The current situation indicates that economics has been thrown into the dialogue about real phenomena that it has not been able to grasp. There are two possible ways out: One is to revert back to the original framework of economics by making the distinction between economics and other disciplines even clearer. The other is to address new difficult questions

^{1 8} Herrmann-Pillath and Boldyrev (2014, Ch.3) attempt at applying their approach to a reconsideration of money, given new historical evidence that money did not emerge to save transaction cost in a barter economy.

submitted inside and outside economics and to use them as a guide for new economic theorizing. As far as I can see, based on the current trend of cross-disciplinary research, economists will follow the latter choice. Then, Aoki's theory of institutions will be recalled as one of the earliest contributions that paved the way.

Conflict of interest statement: The author states that there is no conflict of interest.

References

- Akerlof, G. (1970), "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, Vol. 84, pp. 488-500.
- Aoki, M. (2001), *Toward a Comparative Institutional Analysis*, Cambridge: MIT Press.
- Aoki, M. (2007), "Endogenizing Institutions and Institutional Change," *Journal of Institutional Economics*, Vol. 3, pp. 1-31.
- Aoki, M. (2010), *Corporations in Evolving Diversity: Cognition, Governance and Institutions*, Oxford: Oxford University Press.
- Aoki, M. (2011), "Institutions as Cognitive Media between Strategic Interactions and Individual Beliefs," *Journal of Economic Behavior and Organization*, Vol. 79, pp. 20-34.
- Aoki, M. (2015), "Why is the Equilibrium Notion Essential for a United Institutional Theory? A Friendly Remark on the Article by Hindriks and Guala," *Journal of Institutional Economics*, Vol. 11, pp. 485-488.
- Aumann, R. (1976), "Agreeing to Disagree," *Annals of Statistics*, Vol. 4, pp. 1236-1239.
- Aumann, R. and A. Brandenburger (1995), "Epistemic Conditions for Nash Equilibrium," *Econometrica*, Vol. 63, pp. 1161-1180.
- Clark, A. (1997), *Being There: Putting Brain, Body, and World Together Again*, Cambridge: MIT Press.
- Denzau, A. and D. North (1994), "Shared Mental Models: Ideologies and Institutions," *Kyklos*, Vol. 47, pp. 3-31.
- Fukuyama, F (2011), *Origins of Political Order: From Prehuman Times to the French Revolution* (1st paperback ed.), NY: Farrar, Straus and Giroux.
- Gintis, H. (2009), *The Bounds of Reason*, Princeton: Princeton University Press.
- Greif, A. (1994), "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies," *Journal of Political Economy*, Vol. 102, pp. 912-950.
- Greif, A. and C. Kingston (2011), "Institutions: Rules or Equilibria?" in N. Schofield and G. Caballero (eds.), *Political Economy of Institutions, Democracy and Voting*, Berlin: Springer, pp. 13-43.

- Hayek, F.A. (1952), *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology*, Chicago: The University of Chicago Press.
- Hayek, F.A. (1991), *The Fatal Conceit: The Errors of Socialism*, Chicago: The University of Chicago Press.
- Heath, J. (2008), *Following the Rules: Practical Reasoning and Deontic Constraint*, Oxford: Oxford University Press.
- Hendriks, F. and F. Guala (2015), "Institutions, Rules, and Equilibria: A Unified Theory," *Journal of Institutional Economics*, Vol. 11, pp. 459-480.
- Herrmann-Pillath, C. and I. Boldyrev (2014), *Hegel, Institutions and Economics: Performing the Social*, London: Routledge.
- Hodgson, G. (2015), "On Defining Institutions: Rules versus Equilibria," *Journal of Institutional Economics*, Vol. 11, pp. 497-505.
- Lewis, D. (1969), *Convention: A Philosophical Study*, Cambridge: Harvard University Press.
- McMillan, J. (2002), *Reinventing the Bazaar: A Natural History of Markets*, NY: W.W. Norton & Co.
- North, D. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
- Polanyi, M. (1966), *The Tacit Dimension*. London: Routledge.
- Schotter, A. (1981), *The Economic Theory of Social Institutions*, Cambridge: Cambridge University Press.
- Searle, J. (1995), *The Construction of Social Reality*, London: Allen Lane.
- Searle, J. (1998), *Mind, Language and Society*, New York: Basic Books.
- Searle, J. (2001), *Rationality in Action*, Cambridge: MIT Press.
- Searle, J. (2005), "What is an Institution?" *Journal of Institutional Economics*, Vol. 1, pp. 1-22.
- Searle, J. (2015), "Status Functions and Institutional Facts: Reply to Hindriks and Guala," *Journal of Institutional Economics*, Vol. 11, pp. 507-514.
- Shimojo, S. (2008), *Subliminal Impact* (in Japanese), Tokyo: Chikuma-shinsho.
- Smith, A. (1759/1976), *The Theory of Moral Sentiments*, Oxford: Oxford University Press.
- Smith, A. (1776/1991), *The Wealth of Nations*, Buffalo: Prometheus Book.
- Sugden, R. (2015), "On 'Common Sense Ontology': A Comment on the Paper by Frank Hindriks and Francesco Guala," *Journal of Institutional Economics*, Vol. 11, pp. 489-492.
- Tomasello, M. (2009), *Why We Cooperate*, Cambridge: MIT Press.