Some Consequences of the Game-Theoretic Architecture of the Sciences of Humanity*

Yukio Takahashi

The theory of game originally worked out by John von Neumann, Oskar Morgenstern and John Nash is a breakthrough in the quest for the operative architecture for scientific studies into the interactions among agentive elements in certain specifiable environments, as we find among nations in an international political and economic environment. Thus the theory has now been applied to various fields of science as in social and natural sciences.

As a theoretical linguist, it is just amazing for me to discuss some of the universal properties of ethics. However, I take the present course of time and days to be the most convenient and hence appropriate time to delve into the realm of what is good and bad with respect to something, which should be the principles of ethics.

I have arrived at the notion of “the ethical environment.” I know that the word environment is the key concept to extend my understanding of ethics to the possible application of Game Theory. Thus I am searching for the correspondent notions of agents and payoff in the realm of what is called ethics. I just wonder whether the notion of payoff would be captured in the realm of ethics. No doubt that I would be able to identify what the agents are in the realm of ethics.

I would like to add that the present paper is an assessment of the work on the interdependence on language, evolution, and their ethical function.

1 The Society of Ethical Minds

Applied to the categorization of the various sorts of conscious activities and processes on the one hand and the various intense or weak emotions or what we feel or what we take to be the artistic impressions that cannot easily be verbalized or expressed in linguistic forms, the integrated theory of consciousness whose underpinnings are the principles of Game Theory and the theory of hierarchy of agents as articulated in the concepts of the society of minds would lead us to the assumption that the emotional and the rather subjective aspects of human psychology and the rather logical objective aspects of logical thinking are blatantly successive and essentially only gradual in nature and that the theory language and mind is thus to be evaluated higher if they are adequately organized and formalized as to capture the

* This is an extended version of Chapter Four of Takahashi (2007) A Theory of Multi-Agentive Phonology (DTP Publishing, Tokyo). The research has been supported in part by the Grant-in-Aid for Scientific Research (KAKEN) by The Japan Society for the Promotion of Science (No. 18520389). I would like to thank Susan Unher for her comments on the earlier version of this paper.
Some Consequences of the Architecture of the Science of Humanity

scientifically refutable notion of successiveness of emotion and logicality of thinking. The Platonistic conception of ethics would be re-verbalized as some game-theoretical evaluation of the interdependence of the system of emotionality and the system of logical thinking. Thus, the whole theorization as I here delineated would contribute to the construction of the meta-theory of ethics.

My conviction on the ethical system is that we are witnessing hierarchies of ethical gadgets that define and give certain strategies to live somehow ethically “correct” lives, in which crucial sense of the term strategies ethical systems are also within the realm of the game-theoretic reinterpretation. This is not all surprising, for the theory of game itself is based on the system of values, which in broader terms are verbalized as good or bad. Thus, it is probable that we will arrive at a universal set of principles of ethics, or what is good and justice. This is somehow an overwhelming consequence, which I have to install into the whole framework of my research objective of the Multi-Agentive Phonology, although it is just in the field of theoretical studies of phonological systems. However, this is not a problem.

The high ranked ethical gadget that we may be able to make the target of our mental scrutiny would be the one that evaluate what is the ultimate goal of one's life. The lowest ranked ethical gadget that we may be able to make the target of our scrutiny would be, e.g., how to breathe, which I don't know well.

Just notice the descriptions in the above paragraph define the boundary of our consciousness that directly is related with the ethical gadgets that defines what is good or bad. In this vein, it may be the case that we may somehow witness the homology between the system of ethics and the mental structure of our consciousness.

If we understand that the goal of ethics is to define what is good and bad, or better or worse, then it is probable that the whole system of ethics has a structure that is homological with the mental structure of human being, as is articulated by Minsky. The notable point in this vein is that he postulates feedback relationships between the generative mental gadget and the mental gadget that stores various sorts of information on the outside world.

A theoretical corollary of the combined logical extension would be that the central computational gadget in the ethical system would also have some access to the mental storage or mental lexicon on what lead to good result and on what lead to bad results. If the relationship between the two systems of thought on ethics, that is, the generative ethical gadget and the storage of information on good and bad cannot be strictly correspond to the consequences that are observed in the real world, we may have evoke some kind of inferential system that would search for linkages between the two systems.

The same course of argument would be applied to the system of consciousness, by which I mean that the system of consciousness would have access to the lexicon of what is reality.

One of the major regions in the study of consciousness is the integrity of personality. In a short passage on the game-theoretic homology in the organization of ethical attitudes and language, I referred to the set of information that is stored for in-
spection when the society of gadgets is operated to work out a set of outputs that is eligible for humanely integrated activities.

One of the corollaries of my reasoning is that the psychological phenomenon of what is called the multiple personality disorder (for short, MPD) is a phenomenon in which the highest ranked psychological has somehow gone away from the psychological system so that the second order psychological gadgets, which may be called the sub-personality, are not integrated into a complete whole of personality. The loss of the highest ranked psychological gadget may be suppressed or taken away by some sorts of traumatic events that are very ominous.

The integrity of coherence in the psychological behaviors of a person may be accounted for by alluding to our assumption on the stored memory of who we are, which may be formalized as the functioning of the K-Society in the sense of Minsky. Thus the phenomenon of MPD may also be accounted for as the loss of long-term memory.

In this vein, I am obliged to search for a linguistic counterpart of the MPD, in which a person may lose his/her language due to some traumatic events in life.

2 Ethics and Religion

Developments of the logic toward the meaning of life crucially dichotomize themselves into those based on religion of every persuasion and those that are founded upon ethics. In the realm of ethics, every thread of argumentation leads to, or is reducible into, some definable set of principles of goodness and righteousness, while every persuasion in the religious environments soars into the humanly inaccessible garden of mysticism. I believe that by definition ethics cannot, and should not, be conducive to mysticism.

As far as my understanding is correct, the lines of argumentation in the realm of ethics are traceable with respect to the principles and culture-specific idiosyncrasies, and hence are verifiable according to them.

Dennett (2006) in his talk in TED (Technology, Entertainment and Design) refers to his new concept of “reverse-engineering of religions,” and thus before going on to Dennett’s contention I would like to cite a definition of “reverse-engineering” from a web site in http://www.blkforest.com/:

(1) Reverse engineering (RE) is the process of discovering the technological principles of a device or object or system through an abductive analysis of its structure, function and operation …. It often involves taking something (e.g. a mechanical device, an electronic component, a software program) apart and analyzing its workings in detail, usually to try to make a new device or program that does the same thing without copying anything from the original.

To have a proper understanding of Dennett’s points, we may refer to two types of
Some Consequences of the Architecture of the Science of Humanity

evolution of species, as was propounded in Darwin’s (1909) *Origin of Species*: (i) realization of species by domestication, as we observe the growth of cows from the now extinct aurochs and (ii) realization of species by natural selection, as we find the appearance of *homo sapience* in the biological history.

Dennett (2006) introduces the dichotomy between the religion as a natural phenomenon and the religion “domesticated” so as to promote the purposes of communities. From a game-theoretic viewpoint, the two types of religion have constitutionally different environments: in the latter, religions are evaluated relative to their contributions to the purposes of the communities where they are embraced, while in the former religions are naturally selected by the elements of humanity, one of which is the imaginative ability to arrive at some understanding of the outside world. In this sense, the former are biologically founded.

Going back to the dichotomy of religion and ethics, Armstrong (2005:2) refers to the humanistic foundation of myth and she alludes to the universal traits of myth as follows:

(2) We [human beings—YT] are meaning-seeking creatures. Dogs, as far as we know, do not agonise about the canine condition, worry about the plight of dogs in other parts of the world, or try to see their lives from a different perspective. But human beings fall easily into despair, and from the very beginning we invented stories that enabled us to place our lives in a larger setting, that revealed an underlying pattern, and gave us a sense that, against all the depressing and chaotic evidence to the contrary, life had meaning and value.

In this sense, the growth of myths is a biologically founded process, which is induced by one of the humanly endowed ability, faculty of imagination to arrive at a holistic understanding of the outside world. Thus, Armstrong (1993:3) does not hesitate to affirm:

(3) In the beginning, human beings created a God who was the First Cause of all things and Ruler of heaven and earth.

The link between the notion of religion and the communities and nations has its consequence on the form and content of the religion: The consequence will be accounted for by the pay-off relations in the sense of Game Theory, with which I will be concerned in my paper to come.

One of the ethical virtues, altruism, is a focus of debate in evolutionary biology, as we find in John Maynard-Smith’s “Foreword” and Cronin’s “Preface” to Cronin (1991):

(4) John Maynard-Smith in Foreword

… She [Helena Cronin—YT] has told me much that I did now know about the ideas of Darwin and Wallace, and the disagreements between them. She has also
understood the modern research on the same topics. For Darwin, the ant and the peacock symbolized two major difficulties for his theory—the existence of cooperation, and or apparently maladaptive ornament. I think he would enjoy reading her account of what has happened since.

(5) Helena Cronin in Preface

… Yet historians have largely overlooked both book [The Descent of Man and The Origin—YT] and theory. Equally, little attention has been paid to the history of altruism—above all, to the vague, good-for-the-species view that was rampant from about the 1920s to the 1960s. … The discrepancy between Darwin’s and Wallace’s magnificent legacy and these ungenerous appraisals left me dissatisfied. I decided to explore more of the new territory that Darwinian theory was opening up.

The problem that I would like to take up here is whether the virtue, i.e., altruism, can be a universally given trait of human beings.

From a broader perspective of humanity, Jackendoff (1994:218) summarizes his points as follows:

(6) Let’s now turn to the question that motivates this book: Is there a “human nature”? The picture that has emerged is that our “human nature” consists in having a collection of innate brain specializations or modules, each of which confers on us certain kinds of cognitive powers: the ability to learn a language, to learn to appreciate music, to come to understand the visual world, to construct concrete and abstract thoughts, to learn to function in a social environment, and no doubt more.

It is natural to assume that the innate faculty to learn to function in a social environment has a set of parameters embedded to capture cultural and ethical diversities. No doubt that cooperation, or altruism, contributes to the proper and successful functioning of an individual, or an agent (a player in a game) in a social environment, in which sense altruism is a strategy of an agent that evaluates the proper functioning higher than the self-centered orientation.

Jackendoff (2007:167-71) goes further to define the axioms of cooperation and competition among individuals in groups:

(7) Axioms of Groups
a. Other things being equal, if you are a member of my group, I will behave favorably toward you. In particular, I will be willing to cooperate with you; and I expect the same from you.
b. Other things being equal, if you are not a member of my group, I will behave unfavorably toward you. In particular, I will compete with you; and I expect the same from you.
Jackendoff (2007:169) accepts the notion of the code of conduct, conceptualized as a *joint commitment* of the members, which he dubs *normative principles* (p. 178). What is of game-theoretic interest is that he notices that a “particular action may have conflicting consequences in different normative domains” and that “legal and religious codes can be used to legitimate the raw exercise of power” (p. 180). Remember that Jackendoff (1994) accepts the notion of the innate faculty to learn to function in a social environment. However, he raises four reasons to “avoid the temptation to look for universal morality” (pp. 181-82).

Going back to the problem of universality of altruism and the biological foundation, Wallace (1890:325) cites Darwin’s gloom over the inconvenient fact that “Those who succeed in the race for wealth are by no means the best or the most intelligent, and it is notorious that our population is more largely renewed in each generation from the lower than from the middle and upper classes.” Smith (2008) summarizes Wallace’s points as follows:

(8) ... he rejected the Lamarckian notion that traits acquired by individuals during their lifetimes could be transmitted to their offspring. ... But he also rejected the then popular doctrine of eugenics both on scientific and ethical grounds. ... Freed from the constraint of marrying primarily on grounds of financial security (as Wallace believed capitalist society demanded), women could now choose as marriage partners only those men who possessed “noble” traits such as intelligence, compassion, altruism, and physical health. Conversely, women would reject potential mates men who were idle, selfish, violent, weak in intellect or who possessed any overt hereditary disease.

Wallace (1905:386) confirms his position and interprets it to be the most significant contribution to Darwinism:

(9) … the only method of advance for us, as for the lower animals, is in some form of natural selection, and that the only mode of natural selection that can act alike on physical, mental, and moral qualities will come into play under a social system which gives equal opportunities of culture, training, leisure, and happiness to every individual. This extension of the principle of natural selection as it acts in the animal world generally is, I believe, quite new, and is by far the most important of the new ideas I have given to the world.

Wallace (1890:337) goes on to argue for the “rational social organization” to guarantee “the improvement of the race”:

(10) In the society of the future this defect will be remedied, not by any diminution of our humanity, but by encouraging the activity of a still higher human characteristic—admiration of all that is beautiful and kindly and self-sacrificing, repug-
Yukio Takahashi

nance to all that is selfish, base, or cruel. When we allow ourselves to be guided by reason, justice, and public spirit in our dealings with our fellow-men, and determine to abolish poverty by recognising the equal rights of all the citizens of our common land to an equal share of the wealth which all combine to produce,—when we have thus solved the lesser problem of a rational social organisation adapted to secure the equal well-being of all, then we may safely leave the far greater and deeper problem of the improvement of the race to the cultivated minds and pure instincts of the Women of the Future.

Thus, the “pure instincts” are significant agents in a game-theoretic account of the progress of the human race.

Cronin (1991:257-58) cites Axelrod’s (1984) three properties to arrive at a non-zero sum game to cooperate:

(11) …, being ‘nice’ (never the first to defect), ‘provokable’ (retaining against defection) and ‘forgiving’ (letting bygones be bygones and resuming cooperation). Niceness generates the rewards of cooperation; provokability discourages persistent defection; and forgiveness heads off long, reverberating bouts of recrimination and counter-recrimination. The reason why a strategy with these properties can be so successful is that, while it plays against another such strategy, not least itself, then both players can win the reward for mutual cooperation (R) on every encounter; they can take full advantage of playing a non-zero-sum game to help one another to attain a high average score for each of them.

The notion of “non-zero sum game” is a benchmark for defining the evolution of kinds of processes of change in some specifiable attributes of existence, including the notion of the evolution of language of human beings, as I will try to elucidate in the following section. This argumentation does now always mean, or should not be organized so as to mean, that the several levels of interaction among human beings, animals and insects, and so on, interact with each other to direct themselves to one ultimate optimal state, which is which might be called “predetermined harmony.”

Maynard-Smith (1982) crystallized the notion of “Evolutionarily Stable Strategy,” which itself implies a stable state. However, this is a very simplified formal simulation of the biological environment, which is in reality a consequence of very complex interactions among various types of agents, that is, creatures. I may add that Pinker (2007) contends that violence by human beings is a fractal phenomenon.

3 The Evolution of Language as the Enhancement of the Environment for

---

1 I would like to thank Mr. Yoshiyuki Terui, my colleague, for his pointing out the possibility that the theoretical development of my game-theoretic account of altruism and other matters may be lead to the state of predetermined harmony.
Some Consequences of the Architecture of the Science of Humanity

Non-Zero Sum Games

Chomsky (2003:404) delineates a humanistic view of language based on innateness hypothesis and he goes on to describe the elements of humanity:

(12) Elements of Humanity

A vision of a future social order is in turn based on a concept of human nature. If in fact man is an indefinitely malleable, completely plastic being, with no innate structures of mind and intrinsic needs of cultural or social character, then he is a fit subject for the “shaping of behavior” by the state authority, the corporate manager, the technocrat, or the central committee. Those with some confidence in the human species will hope this is not so and will try to determine the intrinsic human characteristics that provide the framework for intellectual development, the growth of moral consciousness, cultural achievement, and participation in a free community.

The elements are (i) the framework for intellectual development, (ii) the growth of moral consciousness, (iii) cultural achievement and (iv) participation in a free community. They are all founded upon the system of human language.

The former President of the United States of America, Bill Clinton (2000) delivered a speech, in which he praises a book subtitled “The logic of human destiny”:

(13) There is an astonishing new book out, been out a few months, by a man named Robert Wright, called Nonzero—kind of a weird title unless you're familiar with Game Theory. But in Game Theory, a zero-sum game is one where, in order for one person to win, somebody has to lose. A non-zero-sum game is a game in which you can win and the person you're playing with can win, as well. … essentially, as societies grow more and more connected, and we become more interdependent, one with the other, we are forced to find more and more non-zero-sum solutions. That is, ways in which we can all win. … Because we are all part of the same human family, and because, actually, life is more and more a non-zero-sum game, so that the better they do, the better we'll do. (Applause.)

Bill Clinton sees the future of the United States and human beings in Wright’s theses on evolution logically explained.

The evolution of language is an episode that occurred in the process of complication of the biological organization. Wright (2001) refers to the notion of the game-theoretic dichotomy of zero-sum vs. non-zero-sum games to account for the stages of biological evolution:

(14) Evolution as the Enhancement of the Environment of Non-Zero Sum Games

Sometimes political scientists or economists break human interaction down into zero-sum and non-zero-sum components. Occasionally, evolutionary bio-
gists do the same in looking at the way various living systems work. My contention is that, if we want to see what drives the direction of both human history and organic evolution, we should apply this perspective more systematically. Interaction among individual genes, or cells, or animals, among interest groups, or nations, or corporations, can be viewed through the lenses of Game Theory. What follows is a survey of human history, and of organic history, with those lenses in place. My hope is to illuminate a kind of force—the non-zero-sum dynamic—that has crucially shaped the unfolding of life on earth so far.

Thus, if the evolution of human language is a leap from one stage of the biological evolution to another, then language would be argued to be an enhancement of the game-theoretic environments that provide strategies for non-zero-sum situations.  

Thus, going back to Pinker (2007), he cites a “gold rule of thumb” for the evolution of the language of human beings, referring to Wright (2001):

(15) The more you think about and interact with other people, the more you will realize that it is untenable to privilege your interests over theirs, and at least if they do not want to listen to you, you cannot say “My interests are more special compared to yours,” any more than you can say that the particular spot that I’m standing on is a unique part of the universe, because I have to be standing here for minutes.

Thus we may arrive at an understanding of the evolution of human language, which exhibits sets of attributes that are discontinuously different from the communication systems of animals and insects.

---

2 The point that I describe here was made open on the occasion of the research meeting held by the circle of the teachers of English in the Hanamaki area on November 6th, 2007. I would like to thank the organizers of the meeting, especially the president Mr. Kodashima, dean of the East Hanamaki High School.
Some Consequences of the Architecture of the Science of Humanity

References