

# A Preliminary Note for “Bushido AI Programming”

Tadao Onaka

## Contents

1. Introduction: Who’s afraid of Artificial Intelligence?
2. Stewardship: The sole source of social value creativity
3. What is Bushido exploring?
4. What are the value elements of Bushido?
5. Jin: The Benevolence of Bushido
6. Gi: The Justice of Bushido
7. Lei: The Politeness of Bushido
8. Chi: The Wisdom of Bushido
9. Shin: The Sincerity of Bushido
10. The Values from Buddhism and Shintoism

## 1. Introduction: Who’s afraid of Artificial Intelligence?

This paper is a preliminary note for the conceptual model development of the artificial intelligence made of Bushido values.

In 2013, James Barrat warned in his book ‘Our Final Invention’ that the development of artificial intelligence would culminate in its terminating all human beings. However, we can also find an entirely opposite perspective. If you review the development history of cognitive science, one of typical research into artificial intelligence (AI), enhanced by those names such as Marvin L. Minsky of MIT and Herbert A. Simon of Carnegie Mellon, you would find that it contributes, first of all, to analyzing how human beings think and act before developing any robot. Why? It is because, in order to develop any type of AI, beforehand we need to understand how the model human beings think and act, substantially and comprehensively enough to translate the understandings into AI algorithms and programs.

In other words, unless human beings can develop AI which terminates all human beings by its own will, at any moment there exists a risk that human beings themselves accidentally or occasionally terminate the society of human beings before AI they develop would do so. Or positively, if human beings can develop AI which can evolve by itself creatively and sustainably, human beings also can do so before realizing such AI, as human beings need to understand and grasp all the requirements to evolve sustainably, at any rate and by any means, before programming such smart AI.

## 2. Stewardship: The sole source of social value creativity

Then why Bushido values? It is because the most substantial core value of Bushido, the value of stewardship with a sense of individual freedom and independency, seems to be rather quickly diminishing globally. While Bushido is the culture of chivalry traditionally and solely inherited in Japan for about 1000 years, it seems diminishing even in Japan recently. The diminishing of stewardship can be observed from the fact that the value of cultivating own monetary values, the opposite value to stewardship, the desire of individuals, societies and countries to increase their own monetary assets with highest priority, dominates currently and globally among corporations, the largest component of global economy, under the mission of stockholders' property value maximization.

Then, why does the diminishing of the stewardship; the core value of Bushido, matter?

The answer to this question assumes two hypotheses: One is that the stewardship, the value to serve the society, is the sole source for social value creativity. Another is that the diminishing of the creativity of individuals, corporations and global societies is the most critical cause of the current global economic turmoil.

What is the current global economic turmoil?

It is the continuous shrinking of actual economy being depressed, pressurized, and consumed one way by the desire of increasing monetary economy. It is, then, social disparity and consequential violent struggles caused mainly by globally spread immense unemployment. And it is the enlarging gap between actual and financial economies: The enlarging gap further increases the pressure from financial economy to actual economy to forget stewardship but concentrate on pursuing and cultivating various monetary values competitively. Then the diminishing stewardship, the source of creativity, diminishes creativity globally and worsens the turmoil further. This negative cycle of creativity deterioration is another face of the current global economic turmoil.

In summary, it is hypothesized that the most fundamental cause of global economic turmoil is the extreme shortage of the creativity of actual economy compared to the magnitude and the speed of the growth of monetary and financial economy. Also hypothesized is that the stewardship could be one of critical sources for human beings and their corporations to keep themselves creative. Nevertheless, in actuality, so called global standard corporate management that pressurizes human resources to pursue maximizing stockholders' property by overcoming daily survival competitions will not allow human resources to reserve stewardship mind as they are fully devoted to surviving competitions and protecting themselves within merit system rather than to serving others. However, without the stewardship, without the value and mind of serving society, any social value creativity can be hardly realized by nature.

## 3. What is Bushido exploring?

The stewardship, the core value of Bushido, is to pursue the way of living with human dignity that can be realized by serving others and societies with one' own free will beyond one's own cost

and merit sense or scientific rationalization.

Therefore, Bushido has the potential to let us recognize the reality globally that current global societies indulge too excessively in the sense of scientific rationalization raised in the early 20<sup>th</sup> century, indulge in the resulted value to prioritize the cultivation and accumulation of monetary assets, and indulge in the wisdom to make all decisions of individuals, corporations, countries solely based on whether consequences increase monetary assets or not.

Bushido would let us think further: If the indulging dominates globally, the stewardship, the opposite to the value of indulging in cultivating monetary assets, only destined to decline. Then, human societies are only to diminish the source for the creativity to get out of the current global economic turmoil.

The diminishing of stewardship, therefore, would be a serious problem not only because it causes the deterioration of creativity of human beings and worsens the current global chaos but also because it dilutes globally the recognition of human dignity that human beings are the only existence to have the value to serve others as well as to pursue individual freedom and independency.

#### **4. What are the value elements of Bushido?**

In his book 'Bushido' first published in 1899, Inazo Nitobe extracts the elemental values of Bushido. They are five values from Confucianism and some from Buddhism and Shintoism. As Confucianism is basically the modeling of an emperor of an ancient Chinese dynasty, the five values are the requirements for leaders governing nations and organizations. Meantime, the values from Buddhism and Shintoism are the requirements for each individual to live with the celebration by the existence, if we dare to describe, the existence generally called God, which is supposed to embrace the cosmos beyond human beings' imagination. Therefore the values are the requirements to explore endlessly and limitlessly the dignity of human beings in order to appreciate the creation and blessing of the existence embracing the entire cosmos.

The five values of Confucianism are 仁 -Jin (Benevolence), 義 -Gi (Justice), 礼 -Lei (Politeness), 智 -Chi (Wisdom), and 信 -Shin (Sincerity). And if we summarize the values from Buddhism and Shintoism, it can be described as the endless pursuit of the dignity of human beings before the eyes of the existence who observes every thought and action of all human beings. We start examining first the five values of Confucianism and then the value of Buddhism and Shintoism to pursue the dignity of human beings endlessly.

Before starting analyzing each of the five values, it should be noted and understood that the meaning the corresponding English word expresses is quite short and facial compared to the original meaning the Kanji letter (word) carries and expresses. Needless to say, that is why the following explanation of the five values takes pages. Extremely speaking, the corresponding English word cannot translate comprehensively the original meaning. However, for the purpose of sharing them globally, we dare to utilize and rely on the globally common language, English.

Why can't English words correspondingly express the meaning? Because it bears only the

least common meaning of corresponding words of all phonetic languages historically integrated into the current English language, such as old Greek, Latin and French. Moreover needless to say, Chinese and Japanese languages described with pictographic letters are, by nature and definition, not included in the integration.

## 5. Jin: The Benevolence of Bushido

The visual structure of the Kanji of 'Jin' which is generally believed to correspond to the English word of 'Benevolence', means 'human beings with mercy and peace'. It was originally developed in the ancient China around at least 4000 years ago ('Jito', Shizuka Shirakawa). The word and its meaning, then, were escalated by Confucianism emerged in the 6<sup>th</sup> century BC to be the most superior value for the leaders governing nations and organizations. ('Jito', Shizuka Shirakawa)

Why was it perceived as the most superior value? We can find a hint to think of the answer, in the book on the most advanced innovation, artificial intelligence: 'Our Final Invention - Artificial Intelligence and the End of the Human Era'. (Ref. 1)

*As an example of unintended consequences, Oxford University ethicist Nick Bostrom suggests the hypothetical "paper clip maximizer". In Bostrom's scenario, a thoughtlessly programmed superintelligence whose programmed goal is to manufacture paper clips does exactly as it is told without regard to human values. It all goes wrong because it sets about "transforming first of all of earth and then increasing portions of space into paper clip manufacturing facilities." Friendly AI would make only as many clips as was compatible with human values. (Ref. 1)*

What the Friendly AI is expected to have would be one of factors of the Benevolence of Bushido: It can be expressed as the respect to 'all' human values, which can be confirmed by the sentences following the above in 'Our Final Invention'.

*Another tenet of Friendly AI is to avoid dogmatic values. What we consider to be good changes with time, and any AI involved with human well-being will need to stay up to speed. If in its utility function an AI sought to preserve the preferences of most Europeans in 1700 and never upgraded them, in the twenty-first century it might link our happiness and welfare to archaic values like racial inequality and slaveholding, gender inequality, shoes with buckles, and worse. We don't want to lock specific values into Friendly AI. We want a moving scale that evolves with us. (Ref. 1)*

Now we, human beings, in the 21<sup>st</sup> century would have to feel a great irony from the above. Before programming Friendly AI, human beings cannot necessarily respect all values of human beings by being bound by specific dogma or philosophy or religion. As an example, in all advanced societies business persons are bound by the dogma of Global Corporation Management Standard that requires corporations to maximize ROE (Return on Equity). It would be interesting to see the sentences of the previous excerpt making sense naturally, even if we replace the word of 'superintelligence' with 'human beings' and 'manufacture paper clips' with 'maximize ROE' as below:

*A thoughtlessly programmed human beings whose programmed goal is to maximize ROE*

*does exactly as it is told without regard to human values. It all goes wrong because it sets about "transforming first of all earth and then increasing portions of space into ROE maximizing facilities.*

Meantime, Inazo Nitobe (Ref. 3) describes the value of Jin as benevolence but at the same time adds the following explanations to complement the missing: (Jin is) *the feeling of distress, love, magnanimity, affection for others, sympathy and pity, which were ever recognized to be superior virtues, the highest of all the attributes of the human soul.*

Nitobe adds, *He (Mencius) also says that "the feeling of distress is the root of benevolence, therefore, a benevolent man is ever mindful of those who are suffering and distress." Thus Mencius long anticipate Adam Smith who founds his ethical philosophy on Sympathy.*

By that suggestion, we can extract the core essence of Jin or Benevolence of Bushido as 'Empathy' in English from the words of 'mutual sympathy' proposed in 'The Theory of Moral Sentiments' by Adam Smith.

In summary, Jin, the benevolence pursued in Bushido could be 'Empathy with all values of human societies'. Then, it seems unrealistic and impossible to realize it nowadays. However, the realistically impossible attribute of Jin itself is the major reason that Confucius regards it as the highest value. At any rate, whoever intends to develop Friendly AI to prevent non-friendly AI from terminating all human beings for its specifically programmed value or dogma will need to fully understand and master the value of Jin. Then it is also true that until we become able to share the empathy with all values of human societies beyond any specific value or dogma, we are always at risk of terminating ourselves accidentally and coincidentally in pursuing only a specific value or dogma with the sacrifice of others, regardless of whether non-friendly AI would do so or not.

## 6. Gi: The Justice of Bushido

The Kanji of Gi visualizes a goat sacrificed to God and means the pureness and flawlessness of the mind of the server. ('Jito', Shizuka Shirakawa)

Nitobe tries to explain the value of Gi, Justice of Bushido, by spending most of the chapter for the explanation on the value of Giri, which is opposite to Gi and emphasizes that Giri should not to be misunderstood as or confused with Gi. Giri is the value forced by specific culture and virtue of family, organization and community. From his paradoxical explanation, therefore Nitobe points out that Gi, Justice of Bushido, is not based on any specific culture or virtue. Then how or based on what can we judge whether the subject we face accords with Gi or not, and whether it is just or not in terms of Gi?

Michael Sandel in his 2010 book 'Justice' (Ref. 9) discusses what is justice quite comprehensively from the viewpoint of the western culture and history. After having challenged various virtue models from Utilitarianism, Libertarianism, Kantianism and Aristotle's Teleology, he finally concludes that we may share a specific set of value as the base for judging justice only within each community like polis and that it is important for all human beings to keep discussion among different communities and countries on the difference of values, virtues and cultures.

However, if any ancient Confucian would read the book, he or she would question why Sandel's book discusses only the justice for sharing pie fairly among stakeholders. In other words, they are only the discussions to seek for justice on what each stakeholder can get from the others. Is there any other case where the judgment of justice needs to be called for? Yes, a Confucian would also answer. There should exist also the justice on whether or not what you do accords with human dignity blessed by the existence, generally called God: It is the justice on whether or not you prioritize to act on what you can do for others, whether or not you challenge to serve others, rather before to act what you can get for yourself from others.

On this point, Nitobe refers Confucius' words: *"Let but a prince cultivate virtue, people will flock to him; with people will come to him lands; lands will bring forth for him wealth; wealth will give him the benefit of right uses. Virtue is the root, and wealth an outcome"*. Gi, The Justice of Bushido is about whether or not the human dignity of stewardship is prioritized by being liberated from the desire of individual wealth cultivation.

Bushido's justice reminds us of what we have lost if we cannot find any other reason for judging justice rather than the reasons to judge the fairness of sharing various types of pie. Justice exists and is called for also when and where we judge whether or not we prioritize stewardship with human dignity blessed by God than scientific rationalization or the effectiveness of return on investment. If our modern wisdom would be too much occupied by the modern science of rationalization, don't we need to find the complementary wisdom? We discuss what it would be later in Chapter 8. Chi, The wisdom of Bushido.

## 7. Lei: The Politeness of Bushido

The Kanji of 'Lei' visualizes a cup of liquor utilized for ceremony. ('Jito', Shizuka Shirakawa)

While the corresponding English word to Lei would be politeness, again the meaning of the Kanji would be rather different from the English one, which would be inferred from the Confucius definition that Lei is the embodiment of Jin, the empathy for all values of human beings. Confucius strongly warned that it should not be confused with the manners just following facial protocols. Therefore, it would be more closer to Self-Control Nitobe explained in Bushido. It would be translated as the efforts to control oneself to prioritize others' values than one's own. Then, it becomes the source to learn from others and results in more occasions to know others' needs. Then it brings more opportunities to envision creatively if one challenges what one can do for the others' needs however different they are from one's own, which is why Confucius defines Lei as the embodiment or action of Jin.

## 8. Chi: The Wisdom of Bushido

The Kanji of Chi consists of the images of sacred arrow and shield together with a box for commitment letters to God, visualizes the status of knowing. Chi is given by God if one pledges one's words to God. ('Jito', Shizuka Shirakawa)

In Confucianism, learning is defined as not only gathering knowledges but also envisioning what one can do for societies. So the best wisdom is called Ryo-Chi, meaning good wisdom or conscience, which might be similar to the meaning of the English word. The good wisdom is believed to be implemented and to realize good ideas when one thinks with the conscience blessed by God. It would be also the condition that one can control one's desire to dominate with one's own value over the others' and that one can prioritize to envision what one can do for others' values rather than to envision for one's own.

While Nitobe has not established a chapter to explain Chi in his 'Bushido', he explains what prevents warriors from getting to be just savage beasts as below:

*What Christianity has done in Europe toward rousing compassions in the midst of belligerent horrors, love of music and letters has done in Japan. The cultivation of tender feelings, breeds considerate regards for the sufferings of others. Modesty and complaisance, actuated by respect for others' feelings, are at the root of politeness. (Ref. 3)*

The above comment of Nitobe's would need some explanation on the history of verses in Japan. It started from the 8<sup>th</sup> century when a book of people's 4500 verses, 'Man-Yo-Shu' was assembled. And then the other series of books of verses followed. Then, to create verses expressing individual emotion got established as one of education courses for the people ranked as Bushi besides the courses to master Confucianism and military arts and strategies. Consequently Bushi people became capable to observe and express their own emotions and feelings through creating verses, which Nitobe points out as the source for keeping those people's humanity even in the midst of savage battles.

The fact that learning emotions prevented Bushi people from losing humanity and the dignity of human beings would remind us of the coincidental fact that emotional intelligence got proposed by Daniel Goleman in the year 2000 to revive what the intelligence of scientific administration and rationalization for business competition have damaged. It would also suggest that we have creativity always ready to be utilized if we are liberated from the dominating artificial intelligence programs of scientific rationalization.

## 9. Shin: The Sincerity of Bushido

The Kanji of 'Shin' visualizes a human being standing beside the pictograph of word consisting of a pictograph of commitment and a box to contain commitment letter to God. ('Jito', Shizuka Shirakawa)

Mencius left a famous message on what sincerity is: *"I have never seen that any intent with never compromised sincerity is not realized"*. He described the message with 'Shi-Sei', two Kanji letters: One visualizes a sacred arrow and the horizon as the landing target of the arrow. Another is consisting of two pictographs of 'Word' and 'Accomplishment'. Then the two Kanji letters together mean 'extreme sincerity'.

However, no one seems so far to have ever proved the Mencius message to be true, especially from the viewpoint of return on investment. If it were proved, the human society would be filled



with the non-facial or non-ceremonial sincerity. However, judging from the fact that Mencius message has been succeeded and inherited beyond the time of more than 2000 years, we would be allowed to conclude that there should have existed any proof or belief in the message throughout the time period at least in Bushido.

The belief existed in the concept of 'Shi' consisting of two letters of Kanji vertically, Bushi and Mind. Then it means Bushi's vision. There existed a belief that those visions to contribute to societies for their evolution with never giving up sincerity are to be inherited by generations: If one cannot realize what he/she envisions within his/her life with never giving up commitment to serve societies for evolution, it shall be succeeded by next generations until it gets realized. However if we reflect the human beings history of evolution, we find that such belief resides not only in Bushido but also in all human societies.

The fact that only visions to serve societies with never giving up commitment can be succeeded and realized would indicate two facts: One is that it is the learning competency supported by never compromised stewardship vision and commitment that overcomes all problems and barriers until the vision gets realized. The other is that if so, any organization cannot evolve sustainably without such stewardship visions and values. We may accept both by the common sense that the learning competency with the commitment of serving social evolution can only keep the potential to synchronize organizations with social changes.

From this viewpoint, we also come up to find why corporations with human resources and their administrative managements deteriorate by time. The conceptually inorganic existence of human resources and their administrative management would not envision creatively by nature beyond the accountabilities and responsibilities assigned to each individual resource. Those corporations are only destined to be consumed and rust.

For corporations to be sustainable, therefore, they should be organizations of organic existence, human beings, who by nature can envision creatively. Only human beings can creatively envision as long as it is assured that the freedom and independency to pursue their realizing stewardship for the social evolution in accordance with the corporate mission of pursuing sustainability.

Then, we come to understand well now why not a few corporations globally have started shifting their supreme mission from maximizing ROE for being the best property of stockholders, to pursuing the sustainability with environment and society and therefore the sustainability of themselves. ('Captain Planet', by Paul Polman, 'Creating Shared Value' by Michael Porter)

## 10. The Values from Buddhism and Shintoism

Shintoism encourages people to believe that their supreme God always watches everyone. Buddhism then enhances and extends the belief further as such that the representatives of the supreme God resides, in everything in the world, not only in human beings but also in plants, animals and in all natural environments.

By this belief, Bushido's value of Jin, the empathy with all values of human societies, can be enhanced accordingly. And then, we would be encouraged to pursue our reason of existence, the



human dignity of stewardship. And if the pursuit is supported with the self-control to learn other's new needs and to envision what we can do for them, we naturally create new social values.

This chain of thoughts would tell us that empathy and stewardship are the unique blessing to human beings, that we could live with two contradicting values of individual freedom of independency and of serving others, and that it is the human dignity to pursue the way to realize the blessing: If we commit to pursuing the human dignity, we culminate in creating social value in any occasions. And the creativity based on empathy and individual freedom is the embodiment of human dignity.

We are coming up to the era to take off the excess scientific rationalization programmings and decision makings based on just only whether monetary assets to be increased or not consequentially. If we take off such programmings and get engaged in the opposite value to utilize individual freedom for serving others and societies, we realize the dignity of human beings. And that realization of human dignity with the resulted creativity makes possible the sustainable evolutions of human societies. Such chain of thoughts can be proved to be true not only in Bushido but in all societies globally. The historical role and duty of Bushido, coincidentally recorded and described by Inazo Nitobe with his protestant Christianity early in the 20<sup>th</sup> century, turns out to have preserved the essence of the ancient Asian civilizations until the global human society would need it for further evolution.

## Reference

1. *Barrat, James, "Our Final Invention- Artificial Intelligence and the End of Human Era"* 2015 Griffin, Thomas Dunne Books, St. Martin's Press, New York
2. *Goleman, Daniel, "Emotional Intelligence"* 1996, Bantam, New York
3. *Nitobe, Inazo, "Bushido - Samurai Ethics and the soul of Japan"* 1906 by G. P. Putman's Sons, New York, revisions and reprint with revisions in 2004 by Dover Publications, Inc. New York
4. *Onaka, Tadao and Denscombe, Nigel K. , "Corporate Sustainability Governance"* 2015 Global Management Networks Co., Inc. Tokyo, Amazon Japan
5. *Onaka, Tadao, "Shin Kigyo Touchi"* 2015, Global Management Networks Co., Inc. Amazon Japan
6. *Onaka Tadao, "Bushido Management"* 2014, Global Management Networks Co., Inc. Amazon Japan
7. *Polman, Paul and Ignatius, Adi, 'Captain Planet'* 2012, Harvard Business Review, Boston
8. *Porter, Michael, 'Creating Shared Value'* 2011, Harvard Business Review, Boston
9. *Sandel, Michael j., "Justice — What's the Right Thing to Do?"* 2009, Farrar, Straus and Giroux, New York
10. *Shirakawa, Shizuka, "Jito"* 2007, Heibonsha, Tokyo
11. *Shirakawa, Shizuka, "Shoki Man-Yo-Shu"* 2002, Chuou Koron Shinsha, Tokyo
12. *Smith, Adam, "The Theory of Moral Sentiments"*, published originally in 1759, Dover edition, first published in 2006, an an abridged republication of the six edition by A. Millar, 1790.
13. *Yoshida, Shoin, "Kou-Mou Sakki (The Lecture on Mencius Thoughts)"* originally lectured in 1855, published with edition by Keigo Kondo in 1979, Kodansha, Tokyo