Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 60 (2015) 1865 – 1874

Procedia Computer Science

19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

A Note on Continuous Self-Identification as Self-Awareness: An Example of Robot Navigation

Yoshiteru Ishida

Department of Computer Science and Engineering,
Toyohashi University of Technology
Tempaku, Toyohashi 441-8580, Japan

Abstract

This note reports on interdisciplinary approaches to model consciousness, an aspect of self-awareness in particular, aiming at artificial consciousness that can be mounted on an autonomous and mobile robot. For self-awareness to emerge, the self-identification process plays an important role. Self-awareness would emerge when self-locating in a self-created map in robot navigation; when solving self-related problems in (a self-related version of) the *frame problem*; and when a singularity arises in mapping the reference point in mathematical mappings.

Peer-review under responsibility of KES International

Keywords: consciousness model; self-awareness; frame problem; self-identification; robot navigation; fixed point; singularity; free-will

1. Introduction

Consciousness has been discussed in various areas[5,6,9,12-15,17-19] encompassing philosophy, cognitive science, artificial intelligence (AI) and robotics, to name a few. Such discussions have covered diverse issues such as self-awareness, free-will, qualia and so forth. This paper focuses on self-awareness in artificial general intelligence (strong AI) and robotics, and the frame problem in particular, which aims to enable autonomous robots to solve self-related parts in the frame problem.

Studies of consciousness have examined self-awareness[20] and free-will[9] in order to answer what they are; whether they can emerge even for machines (computers)[12]; and how they can be implemented on computers or robots. This note focuses on the problem of how self-awareness may be characterized in three domains: the robot navigation (robot self-locating) problem in robotics; the frame problem (self-modeling problem) in AI; and fixed points and singularity (in a self-mapping) in mathematics.

These three views make it possible to explain oneness, invariance, and free-will, respectively. Another important property of consciousness is a feeling of time flow, which cannot emerge from a logical inference system with binary values of true or false. Imaginary values are also required to introduce oscillation and thus allow a feeling of time. Consciousness will be modeled as a majority network which has a mode of oscillation.

Nomenclature

$\boldsymbol{x}(t) = (x_1(t), \ldots, x_N(t))$	the state of nodes 1, 2, …, N of a network at time t
$\{\boldsymbol{x}_s: f(\boldsymbol{x}_s) = \boldsymbol{x}_s\}$	fixed points of a function $f(\boldsymbol{x})$

* Corresponding author. Tel.: +81-532-44-6895; fax: +81-532-44-6895.
E-mail address: ishida@cs.tut.ac.jp

1877-0509
Peer-review under responsibility of KES International
doi:10.1016/j.procs.2015.08.297

We suppose that self-awareness emerges as a result of continuous self-imaging and self-identification from sensory inputs in animals which are autonomous and mobile agents. We also conjecture that self-awareness has a merit of increasing the survivability of mobile agents and has evolved for intelligent mobile agents such as humans. Humans have a world model in reasoning many things, and self-awareness plays a critical role for self-related problems as mobile agents. Moving the (self) body to a place is an example of self-related problems as a mobile agent.

For realizing some aspects (such as self-awareness) of consciousness as an operating system (OS) for autonomous and mobile robots, not only philosophical discussions in AI but also engineering discussions in robot navigation and formal discussions in mathematics may be needed. In order to explain the hypothesis that self-awareness may be a continuous self-identification, our approach involves three distinct fields: robot navigation in robotics; world and self-modeling in AI; and fixed points and singularity in mathematics. Each of these three fields has its own role in this study: robot navigation to give examples; AI to set problems; and mathematics to formalize.

To fully utilize the parallelism among these three fields, Sections 2, 3 and 4 are organized in parallel. Section 2 characterizes self-awareness using robot navigation problems. Section 3 characterizes self-awareness in self-related problem solving using a frame problem. Section 4 characterizes self-awareness as singularity of self-mapping. Section 5 discusses how other aspects of consciousness may be related using the characterization of self-awareness. Section 6 suggests the difficulty of realizing free-will with machines and leaves the characterization and formalization of free-will as future challenges.

2. Robot Navigation Problem as an Extendable Component for Robot OS with Self-Awareness

2.1. Self-awareness is "for-by-of" self-identification

Mobile robot navigation can be used to explain the consciousness emerging from the self-identification process, since an important function of consciousness is also to navigate the self as a mobile agent in a physical space. Indeed, both self-localization (self-identification in a physical space) and building a map (world modeling) in which the location is identified are critical in robot navigation as well as in the act of consciousness.

Mapping from sensory inputs to internal images may be captured by the concept of fixed points with an updating function, or invariant sets in dynamical systems. Dynamical systems have been widely used to characterize consciousness.[1,6] Consider a networked automaton expressed as a network. Each automaton (expressed as a node) can take inputs from the external world or from the outputs of other automata, and can feed its output to other automata. Let $\boldsymbol{x} = (x_1, \ldots, x_N)$ be the sensory input to nodes 1, 2, …, N. We denote $\boldsymbol{x}(t) = (x_1(t), \ldots, x_N(t))$ as the state of nodes 1, 2, …, N at time t. The state of nodes is updated by a function $f(\boldsymbol{x})$ and the fixed points $\boldsymbol{x_s}$ that satisfy $f(\boldsymbol{x_s}) = \boldsymbol{x_s}$ are the mapped image of the sensory input. As an example of a networked automaton, the majority network is explained in Example 6.

2.2. Robotic mapping: a primitive toward a robot OS realizing self-awareness

To focus on the robot navigation problem, we assume that the feature of the world objects of interest is only their position and their inter-relation (relative positions).

Example 1. (Positioning a target object: a robot with a sensor**)**

The simplest example to locate a target object with a sensor is: a robot mounted with a sensor that can detect whether the target object exists within the area of a circle with a constant radius. If the sensor on the robot cannot narrow down the target position, the robot moves and tries again (Fig. 1). The target object can be positioned relative to the robot, rather like drawing a Venn diagram in set theory and finding the intersection of the circle areas which the sensor detected.

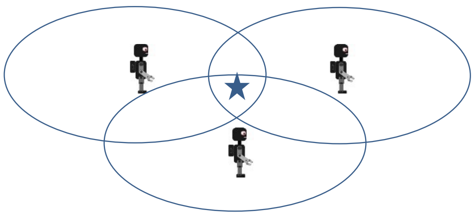

Fig. 1. A robot mounted with a sensor that can detect a target object (star mark) in a circle centered on the robot.

By locating landmark objects and identifying their relative positions, a map for the robot can be created. Current robot technology uses a sophisticated map such as those involving probabilistic representations.

2.3. Robotic self-locating: another primitive toward a robot OS realizing self-awareness

Let us proceed to a self-related problem: self-locating in the map. Although self-locating can be done precisely with a global map such as the latitude-longitude coordinate system of the Global Positioning System (GPS) using GPS satellites, the following example uses an old navigator to clarify what we mean by self-discrimination and self-identification within the context of self-locating.

Example 2. (Positioning itself: a robot with an old navigator**)**
Suppose a robot with an old navigator wants to locate itself. The robot can know all the robots in the vicinity as dots on a map in real-time, but how can it know which dot is itself among these three dots? If a robot moves differently from other robots for self-discrimination (e.g., circling) (Fig. 2), then the robot can know that the circling dot is itself, which we call self-identification.

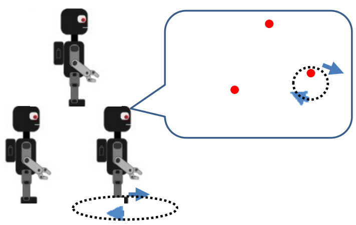

Fig. 2. A robot with a navigator that can locate all the robots in the vicinity as dots on a map in real-time. A robot can identify which dot is itself by moving differently from other robots (e.g., circling).

2.4. Self-awareness is online real-time self-locating in a self-created map with self-mounted sensors

Self-awareness restricted to self-locating in the world map comes from continuous and real-time self-identification based on sensors matching the real (true) self-position felt by the robot itself. Suppose a robot has not only sensors for positioning external objects (and an intelligent mechanism that not only relates these external objects (to create a map) but also relates these external objects with itself (to locate itself within the map)), but also sensors for monitoring the internal state of itself such as the orientation of the robot and for how long the robot moved. In this case, the robot can solve the self-locating problem in two ways: either by creating a map and locating itself within the map by the sensors for positioning external objects; or by monitoring itself and locating itself relative to the initial position. The robot can self-identify continuously by doing online and real-time matching between the two self-positions sensed in these two ways.

We suggest that self-related problems (with self-monitoring sensors) are important for designing self-aware robots with artificial consciousness since humans as ultimate self-related problem solvers maintain their identity by continuous self-discrimination and self-identification with the immune system, which could be a driving force for next-generation AI. The first step to developing robots with artificial consciousness is to deal with the self-related problems.

3. Frame Problem to Place Self-Awareness in Self-Related Problem Solving

3.1. Self-awareness is "for-by-of" self-management in self-related problem solving

A simple way of assessing the intelligence of an animal is to place a mirror in front of the animal. If the animal identifies its self-image in the mirror as himself/herself, then we can assume that the animal has a certain level of intelligence, i.e., a self-related problem-solving capability enabled not only by world modeling but also by self-identification in the world model. We use the frame problem[2] as a starting point to remind ourselves that self-related problem solving is imperative for AI to deal with the frame problem. The point is that not AI with a conventional problem solver but AI involving the self-related problem solver is imperative, which can be a first step toward AI that has self-awareness or even consciousness.

Problem 1. (A frame problem: a self-related analogical reasoning**)**

Suppose a fire breaks out in a house. The entire family except for one baby escapes from the house. An intelligent robot is asked to fetch the baby. The robot has AI that includes not only a general problem solver but also a self-related problem solver that enables the robot to move to a position from where it can extend its arms, hold the baby without crushing it, and even carry the baby with a posture that allows the baby to breathe, etc. However, since the robot also has a general problem solver with the important heuristic of separating the target object into components which can be held and carried more efficiently and independently, the robot fetches the head, body, arms, and legs of the baby independently.

With this frame problem modified to focus on self-related problem solving, it is important to consider not only self-related analogical reasoning (compassion), but also the management function of consciousness as well as expanding the world model to include the self, which we want to define as the act of free-will.

3.2. World modeling: a primitive to deal with self-related problems

To deal with the frame problem, robots need to be able to recognize objects and relate the recognized objects in world modeling. The following example illustrates that even humans can experience visual illusion if the mapping in object recognition involves redundant mappings.

Example 3. (*Necker Cube:* an illusion in object recognition**)**[7]

Although the cube consists of 12 line segments, we focus on six of them (we could alternatively use six faces). These six segments are numbered as shown in Fig. 3[7]. These numbers also indicate the numbers (labels) of the nodes in a sensory part (left-side nodes in the bipartite graph shown in Fig. 3). When node *i* in the sensory part is paired to node *j* f (b) in the recognition part (right-side in the bipartite graph shown in Fig. 3), then the automaton understands that bar *i* is in the front (behind). Thus, the two mappings create two understandings of the cube in the assignment shown in Fig. 3.

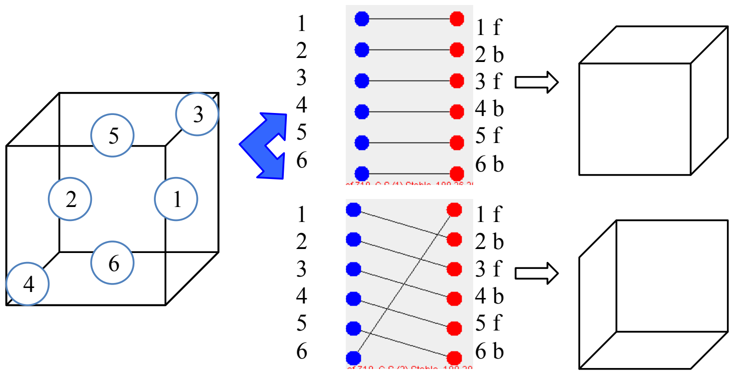

Fig. 3. Perceptual switching experienced when observing the *Necker Cube*. A cube whose six line segments are numbered (*left*). Two matchings showing two possible interpretations (*middle*). Two cubes corresponding two understandings to the matchings (*right*).

It is interesting that the matching mechanism used for object recognition may be used for relating recognized objects in the world modeling. Although object recognition may be realized in a discrete space in the above example, it can also be realized as a stable equilibrium in a continuous space.

3.3. Relating the self: another primitive to deal with self-related problems

As an example of relating the self in the self-created world model, we use a simplified one where the world model is a one-dimensional network where each node is labeled so that the self can be readily related. The following example again concerns the robot self-locating problem, and is indirectly (through the book by Diaconis and Graham[2]) based on Sinden's idea that labeling the path under the robot with a *de Bruijn* sequence allows the robot to locate itself on the path.

Example 4. (Object recognition when the object is the self**)**

Suppose a robot lives in a one-dimensional network world where each node is labeled by a letter, and hence its world model is also a one-dimensional network, which can be trivially realized by copying the world to its memory as a one-dimensional network with the label. Also suppose that the only concern for the robot is the self-location problem. If the robot has a sensor that can read letters in the neighborhood of its position (three consecutive letters: the letter where the robot is located and one node ahead and one node behind in case of Sinden's idea), then the robot can self-locate in its world model.

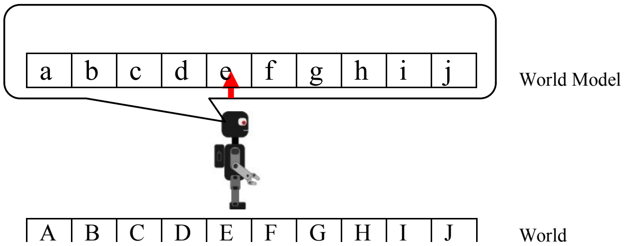

Fig. 4. A robot living in a one-dimensional world where each place is labeled by a letter can self-locate itself in its world model (indicated by a red arrow) by sensing and identifying the letters in the neighborhood of its position.

In this example, the world and the world model are identical. Sinden proposed that a labeling on the path by a *de Bruijn* sequence allows the robot to self-locate on the path[4].

Since self-related problems (self-locating problem in this example) can be a critical problem for animals, it may be important to self-identify through several sensors with distinct modes (e.g., sight, hearing, touch, smell), and to check the results of the self-identification on their cognitive map with consistency against itself.

3.4. Self-awareness is recognizing itself in the self-related problem solving

For a robot with (strong) AI to deal with self-related problems as found in the frame problem, the AI in the robot requires a self-related problem solver. To determine whether we can call emergent phenomena (if they were to emerge) self-awareness or not, we must ask the robot how it feels, or perhaps ask the robot to define the synchronized mode (what is identified with the real self) as self-awareness.

It should also be noted that humans (as sophisticated self-related problem solvers) are always dealing with self-related problems. Writing, submitting and (if accepted) publishing this note itself is an act of attempting to relate the self (this note on self-awareness) to the world model (a literature network on the consciousness), to solve the self-related problem.

4. Fixed Points and Singularity to Formalize Self-Awareness

4.1. Self-awareness is "for-by-of" singularity in fixed points in a mapping

An important property of the (globally stable) fixed point is that it can start from an arbitrary state, and the image (result) is fed again and again.

When you first bought an electronic calculator, you might have enjoyed some mathematical experiments: if the calculator has a button for calculating a specific function, say the square root of an input value, you first input any value larger than one and calculate its square root. Then you input this result and calculate its square root. If you continue this procedure many times, the value becomes closer to 1, which is one of the fixed points of the square root function: $f(x) = \sqrt{x}$. Even this elementary arithmetic experiment indicates two important things: the value (fixed point) obtained does not depend on the initial input value (provided that it is not smaller than 1); and this fixed point experiment can be done not only for the square root function, but for many other functions that make the given value (greater than 1) smaller than the given value but still greater than 1; or for functions that make the given value (smaller than 1) greater than the given value but still smaller than 1.

The self-referential paradox is also attributed to fixed points of functions. Consider the sentence: "this sentence is false." If we believe this sentence is true, then it must be false as asserted, while if we believe otherwise then it must be true; thus the logical value oscillates. That is $x_1 = f(x_2)$; x_2 refers to x_1 to determine x_2, that is, $x_2 = f(x_1)$, which together lead to $x_1 = f(f(x_1))$ and $x_2 = f(f(x_2))$. We can say that both x_1 and x_2 are fixed points of the double-folded function $f(f(x))$; or that the set $\{x_1, x_2\}$ is an invariant set of the function $f(x)$. The following example is also the same case.

Example 5. (A fixed point expressed as an imaginary value is oscillation as time involved behavior**)**
Let us use the function: $f(x) = -\frac{1}{x}$. If we try to determine the value of $x = f(x)$ and assign the value $x = 1$, then $x = -1$, however assigning $x = -1$ would make $x = 1$. This oscillation is also due to the self-reference: using x itself to determine the value of x.

It is well known that the latter case of mathematical self-reference of a function can be resolved by involving an imaginary value, however the former case of logical self-reference can be consistently explained by introducing an imaginary value[10]. Imaginary values in logic may also be related in quantum computation as "square root of negation".[4]

The introduction of imaginary values in the real world in effect means the introduction of a new dimension in the real world: the time dimension. That is, since we cannot determine between mutually excluding values $x = 1$ and $x = -1$ (true or false in logic), we introduce time dimension $x(t) = 1$ and $x(t + 1) = -1$. Indeed, oscillation would happen instead of the logical paradox and value-determining deadlock as noted in Section 6. Time flow is a result of the imaginary value that adds the time dimension to logical values.

The majority network is a networked automaton whose state at time t is $\boldsymbol{x}(t) = (x_1(t), \ldots, x_N(t))$ which is updated by the majority function $\boldsymbol{x}(t + 1) = \boldsymbol{f}(\boldsymbol{x}(t))$ where $x_i(t)$ represents the state of node i. The majority function determines its next state as the majority state of the nodes adjacent to the node. For tiebreaks, if two states requiring a tiebreak are not the same state as the current state of the automaton, then randomly choose either of the two. If one of the two states requiring a tiebreak is the same as the current state of the automaton, then the automaton remains in that state in the next step, too.

Example 6. (Majority network expressed as a generalized bipartite graph**)**[8]

The majority network expressed as a generalized bipartite graph is shown in Fig. 5[8]. A generalized bipartite graph consists of two parts such that each node in the graph must satisfy the condition that the *o-degree* is greater than the *i-degree* where *inner-degree* (*i-degree*) is the number of edges with the nodes in the same part, and *outer-degree* (*o-degree*) is that with the nodes in the other part.

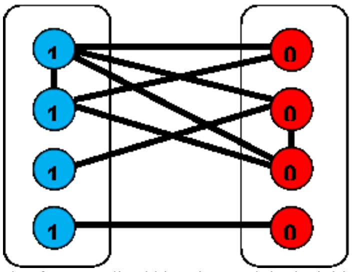

Fig. 5. An example of a generalized bipartite graph in the initial condition, after which the state will oscillate (attractor with period two). The state is indicated by the color (blue and red) as well as by the number (1 and 0), respectively.

Majority networks have several interesting properties; we focus here on their stability. Any of the majority networks defined above will converge on either a fixed point (attractor with period one) or an alternating state (attractor with period two), regardless of the graphical structure and the number of states allowed. The existence of a fixed point $(x_s, x_s, \ldots, x_s)$ is trivial by the definition of the majority network. The existence of an alternating state $(x_1, \ldots, x_1; x_2, \ldots, x_2)$ and $(x_2, \ldots, x_2; x_1, \ldots, x_1)$ can be also understood when the network is drawn as a generalized bipartite graph (Fig. 5) where the states of the nodes belonging to the left part are specified by the first M elements before the semicolon and those for the right part by the last N elements after the semicolon: $(x_1(t), \ldots, x_M(t); x_{M+1}(t), \ldots, x_{M+N}(t))$.

4.2. Defining a map: a primitive to formalize self-awareness

To focus on the robot navigation problem, we assume that the feature of the world objects of interest is only their position and their relation among the world objects (relative positions).

Example 7. (Graph isomorphism**)**

A graph is a mathematical structure consisting of a set of nodes and a set of edges connecting the nodes. For one graph $G(N_G, E_G)$ to be isomorphic to another graph $H(N_H, E_H)$, not only the set of nodes N_G to be mapped to N_H with a bijection but also the connecting structure must be preserved, i.e., when two nodes are (not) connected in G their mapped nodes must be (not) connected in *H* (Fig. 6).

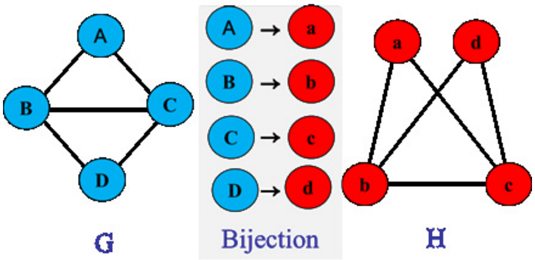

Fig. 6. The graph G on the left is isomorphic to the graph H on the right. The bijection from the set of nodes in the left graph to those in the right graph is indicated in the middle by arrows. The one-to-one correspondence is also indicated by the same letter. If two nodes are (not) adjacent in the left graph, their mapped nodes must (not) be adjacent in the right graph to be isomorphic.

As illustrated by the above Example 7, in defining a map from the world to the world model, not only must objects be recognized and mapped to the counterparts in the model, but also the relation among the objects must be preserved from the world to the world model. We note that graph isomorphism can be done in a dual manner: first map the set of edges and then check the edge connection to the nodes. This would suggest that object recognition as well as object relationing may be done by the same mechanism: fixed point identification in the mapping.

Graph isomorphism is known to belong to NP but not yet even associated to NP-complete. This would suggest that matching the objects and their relations between the world and the world model is computationally intensive with the currently known algorithms.

4.3. Recognizing singularity: another primitive to formalize self-awareness

In the mathematical formalization of self-awareness, one possible way is to involve singularity in mappings. The following example of a Riemann sphere illustrates singularity.

Example 8. (Riemann sphere**)**

In mapping from a point on a sphere to the plane cutting the Equator (regarding the sphere as the Earth) of the sphere as illustrated in Fig. 7, mapping can be done by drawing a line from the North Pole (the reference point) to the plane. Although this approach maps successfully almost all the points on the sphere to the plane, a singularity arises when we try to map the North Pole (the reference point itself).

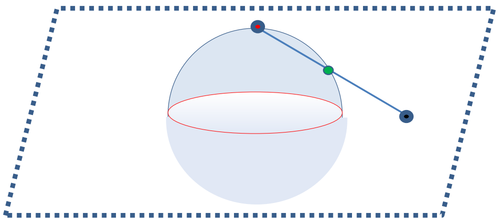

Fig. 7. Mapping from a point on a sphere (green dot) to a point on a plane (black dot) cutting through the Equator of the sphere (red circle) requires the North Pole (red dot) to be referenced. A singularity occurs when mapping the reference point itself.

We assume that a *singular feeling* of self-awareness comes from singularity similar to that defined in the above Example 8, for the self acts as a reference point in the (cognitive) world modeling, and also the recognizing agent and the object being recognized must be equated.

4.4. Self-awareness is a singularity in a mapping of the world modeling

Let us regard the world modeling as a map from the world to the world model. Then the map can be regarded as an automorphism when the world modeling is successful enough to regard the world model as homomorphic to the world, and we can conceive fixed points in the automorphism (the map from the one mathematical structure to itself). Object recognition can be regarded as fixed points in the automorphism (time parameterized function from a set to the set). For the world modeling to be complete, the mapping should involve checking the consistency of relationships as conceived from the graph automorphism example (Example 7).

To characterize self-awareness as a singularity, the self should be placed as an object in the above world mapping. The self can be characterized as a fixed point (the self is one object to be mapped in the world modeling), but the significance of the self is the singularity that the self makes, for the self is used as a reference for relating objects (e.g., their coordinates can be relative to the self) in the world modeling. The singularity may be understood by analogy to the singularity in the Riemann's mapping from a sphere to a plane (Example 8).

Although we characterized self-awareness as singularity of fixed points, the fixed points of double-folded character of the operator and operands may be attributed to the fixed points of $f(f) = f$ whose solution depends on the mathematical structure of the function space itself. When self-awareness is characterized as double-folded fixed points of $f(f) = f$, it requires not only automorphism but also the set of functions must be equal to the set to be mapped (domain) and the set of being mapped (co-domain). We consider that DNA self-replication is an example of this type of fixed points when DNA works as an operator and an operand resulting in the copy of itself.

5. Discussion

Since consciousness has many aspects (not only self-awareness), it is highly likely that other aspects such as free-will may be required to deal with self-related problems as found in the frame problem. However, several phenomena may be tentatively explained from the above characterization of self-awareness.

Oneness comes from that the self (as an object in the world) is identified continuously as oneself (as a mapped object in the world model). Continuity (or feeling of time-flow) comes from the fact that this self-identification is successfully and unconsciously done throughout time. When this self-identification is not successful, strange things will happen to oneself, as is experienced while dreaming when asleep. This mis-identification of the self occurs due to the insufficiency of sensed data because visual sensing is not working, even though auditory sensing can work from time to time, which could cause framing of the strange stories in dreams. We suppose this self-identification process is always working unconsciously, and the unconscious process may be made conscious when the self-identification fails to work properly; this can be experienced as motion sickness when subjected to extreme motion (e.g., in a rotating cup at a fairground). Not only self-identification but also object recognition is a continuous and unconscious process, for we can experience optical illusions such as the Necker Cube and Rubin's base.

It can be imagined that emotions play a certain role in self-related problem solving. Compassion is explained as a self-related analogical reasoning (Problem 1). Anger can be assigned a role in the self-related problem of competition involving the self (including fight). Sadness may play a role in preserving the self and offspring (the self genes). It is interesting that compassion is often associated with sadness; this could be explained by the fact that the associations emerge in an attempt to preserve self-similar objects such as pet animals in addition to offspring.

This note is limited to the discussion of consciousness inferred from several properties such as oneness, continuity, time-flow, free-will, and qualia. These properties derive from the fact that humans are animals, which are moving and self-surviving agents. It is tempting to define consciousness in a general fashion that could encompass entire animals and even plants, however, this note is limited to humans, and so focuses on the consciousness of humans as an animal species.

A motivation of the frame problem (Problem 1) is to examine the significance of self-related problems, and that consciousness and its character of self-awareness and free-will play a significant role in solving self-related problems. Without self-related problem solving, a robot cannot even move to a position from where it can extend its arms, or even if the robot does reach the position, it cannot hold a baby without crushing it, or even if the robot can gently hold the baby it may not be able to carry the baby with a posture that allows it to breathe.

Another motivation for the above paradox is that solving the self-related problem may be important for problem solving: analogical reasoning applied to the self. Analogical reasoning is an important component of AI, for it can be applied to events and entities that have a similar structure. When analogical reasoning is applied to the self, such feelings as compassion emerge. The significance of compassion in solving self-related problems may not require further explanation than the modified version of the frame problem (Problem 1). It is also possible that *developmental disorders* found even in adults may be the result of components of the brain in charge of self-related modeling and/or solving. Modeling of consciousness would contribute to understanding several emotions and also should be constructed on the basis of emotions.

6. A Challenge: Free-will and Robot Mutual Avoiding Paradox

Free-will, which is a significant aspect of consciousness, plays an important role in self-related problem solving. Free-will allows a problem-solving agent to choose options *freely* without any constraints. To suggest its difficulty in logic, we modify an idiom derived from Chinese literature of Han Feizi into a paradox (Paradox 1) in a robot navigation context. The idiom is MU-JYUN in Japanese where MU means a pike and JYUN means a shield, suggesting a contradiction when both a complete pike (that can pierce any shield) and a complete shield (that can protect from any pike) exist. The Japanese word MU-JYUN, which means contradiction, originates from this idiom.

Paradox 1. (Robot mutual avoiding paradox**)**
Two robots, Alice and Bob, are traveling in opposite directions when they meet in the middle of a one-dimensional road with two lanes, facing and trying to avoid each other (Fig. 8). Alice has AI that can predict the behavior of other robots, while Bob has AI that can respond to the contrary of any prediction by other robots. In the simplest version, Alice has an automaton that can output Right (Left) when Bob's automaton outputs Right (Left); and Bob has an automaton that can output Right (Left) when Alice's automaton outputs Left (Right).

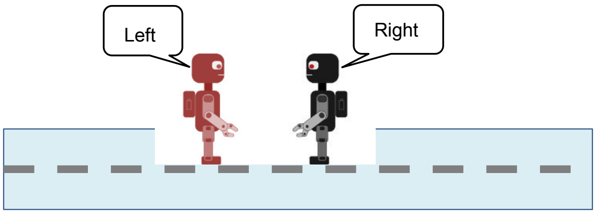

Fig. 8. Two robots: Alice (left) and Bob (right) with two omnipotent AI: Alice can predict any behavior of other robots, while Bob can respond to the contrary of any prediction by other robots. They are deadlocked in a one-dimensional world with two lanes.

This paradox aims to show the significance of symmetry-breaking when the problem is fixed to symmetric situations that hamper the problem to be solved. If one of the agents in the above paradox is human, then the human can recognize the symmetric situation and will break the symmetry by asking the robot to remain still for a moment. We suppose the two components of symmetry recognition and symmetry breaking in self-related problem solving are significant acts of consciousness, namely self-awareness and free-will.

One may argue that the above symmetry recognition and breaking may be solved without considering the act of consciousness by explicitly implementing a protocol (e.g., right lane only) for both robots. But how can they come up with the protocol and how can they negotiate to make the protocol common? There are, however, many more self-related problems even when restricted to the simple paradox. The point is that they are not enumerated and described beforehand, so self-related problems cannot be solved in principle. Logically, this is similar to the fact that one cannot prove problems by giving examples, with the exception of mathematical induction where the problems have symmetry with which entire problems of interest will be systematically enumerated. It seems to be impossible to implement free-will on any automaton by definition: free-will must be free from any constraints on choosing options, whereas an automaton determines its behavior (outputs and state transition) depending on constraints (state-transition rule). If we attribute symmetry-breaking (expanding options or expanding the world model freely) to free-will, it would be even harder for any automaton to realize.

There is a solution, however, to avoid this mutual deadlock. Let us return to the paradox in terms of time. If there is no delay time at all from sensing and decision-making to actions, then Alice and Bob are in a complete deadlock standing still and facing each other. If we allow the inclusion of a delay time for Alice or Bob but the time is too small, then Alice and Bob oscillate with periodic behavior: RLRLRL … (in their actions or their automaton outputs). However, if the delay time is long enough to cover one period of the oscillation then they can successfully break the symmetry without resorting to the act of free-will (or consciousness). Such deadlock and oscillation are parallel counterparts to solving the imaginary value and oscillated behavior of the function in Example 5. It is suggested that free-will is difficult to attain artificially, but we can seek a possible solution in symmetry-breaking by unavoidably and unintentionally including delay times in robots in the real world (indeed, the immune system actively uses errors). In a global design of a collection of robots, heterogeneous robots may be a solution, just as the immune system actively uses diversification.

Since these self-related problems are often conflicting and have trade-offs, consciousness must play a role in managing and reconciling them. Emotions may play a role in (or result from) the reconciliation process in self-related problem solving.

Formalizing free-will and characterizing it as a problem of automatons, whether it exists or not[1,9,11,16], similarly to the halting problem of the Turing machine, is a challenge not only for strong AI but also for autonomous robotics and even for singular points and fixed points of the world modeling as a mapping.

7. Conclusions

We have characterized self-awareness (an aspect of consciousness) as self-identification. The characterization incorporates parallel arguments from three problems in each of three domains: navigation in robotics, the frame problem in artificial intelligence (AI) and singularity of fixed points in mathematics. The aspect of self-awareness alone cannot deal with the self-related version of the frame problem. Aspects of free-will such as choosing options freely and expanding the options by reframing the world model are needed, but their characterization is another challenge.

Acknowledgement

Unfortunately this note may have failed in a self-related problem (in the context of writing a paper): finding already known and published knowledge; relating them for an academic map of this domain; and relating the self (this note) within the map.

References

1. Balduzzi D, Tononi G. Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS Comput Biol* 2008; **4**:6 e1000091.
2. Dennett D. Cognitive Wheels: The Frame Problem of AI. *The Philosophy of Artificial Intelligence* 1984; 147–170.
3. Deutsch D. Quantum Computation. *Physics World* 1992; **5**: 57-61.
4. Diaconis P, Graham R. *Magical mathematics: the mathematical ideas that animate great magic tricks*. Princeton University Press; 2012.
5. Edelman GM. *Wider Than the Sky: The Phenomenal Gift of Consciousness*. Yale University Press; 2005.
6. Freeman WJ. *How Brains Make Up Their Minds*. Weidenfeld & Nicolson: London; 1999.
7. Ishida Y, Yamanaka Y. A note on mapping a matching problem to a majority network as a Lyapunov function: problem solving by synchronization. *Procedia Computer Science* 2014; **35**: 1370–1378.
8. Ishida Y, Hayashi T. Asymmetric phenomena of segregation and integration in biological systems: a matching automaton. *Lecture Notes in Computer Science* 2009; **5712**: 789–796.
9. Kane R. editor. *The Oxford Handbook of Free Will*. Oxford University Press; 2011.
10. Kauffman LH. Self-reference and recursive forms, *J Social Biol Struct* 1987; **10**: 53–72.
11. Laureys S. Eyes Open, Brain Shut. New brain-imaging techniques are giving researchers a better understanding of patients in the vegetative state. *Scientific American* May 2007; http://www.scientificamerican.com/article/eyes-open-brain-shut/.
12. Koch C, Tononi G. Can Machines Be Conscious? *IEEE Spectrum* June 2008; **45**:6 55–59.
13. Koch C, Tononi G. A Test for Consciousness. *Scientific American* June 2011.
14. Koch C. *The Quest for Consciousness: A Neurobiological Approach*. Roberts & Company Publishers; 2004.
15. Laureys S. (ed) *The Boundaries of Consciousness: Neurobiology and Neuropathology* 2006; **150**, Elsevier; 2006.
16. Maye A, Hsieh C-H, Sugihara G, Brembs B. Order in Spontaneous Behavior. *PLoS ONE* 2007; **2**:5 e443.
17. Tononi G. Edelman GM. Consciousness and Complexity. *Science* 1998; **282**: 1846–1851.
18. Tononi G. Consciousness as Integrated Information: a Provisional Manifesto. *Biol Bull* December 2008; **215**: 216–242.
19. Tononi G. An information integration theory of consciousness. *BMC Neuroscience* 2004; **5**:42.
20. Wolf FA. On the quantum mechanics of dreams and the emergence of self-awareness. In Hameroff SR et al. editors. *Toward a Science of Consciousness*. MIT Press; 1996. pp. 451–467.