15TH NANZAN INSTITUTE SYMPOSIUM

Brain Science and Religion

Some Asian Perspectives

Alena Govorounova

Nanzan Institute for Religion and Culture

The Nanzan "Global Perspectives on Science and Spirituality" (GPSS) project sponsored its first international workshop in Taiwan in collaboration with the Center for the Study of Science and Religion at Fu Jen Catholic University in October 2007, on the theme "Consciousness, Brain Science, and Religion." A second international workshop was held in October 2008 in collaboration with the Sogang University Institute for the Study of Religion in Seoul, Korea (see last year's Bulletin 33: 22–56). A final, comprehensive international symposium brought together participants from the earlier workshops as well as other prominent scholars from around the world: Japan, Korea, Hong Kong, Taiwan, India, Russia, France, Canada, and the United States. The following summary was prepared by Alena Govorounova, who reported on the earlier Korea workshop and attended the final symposium.

An international conference titled "Brain Science and Religion: Some Asian Perspectives" (27–29 November 2010) was the closing event of the five-year long GPSS project in East Asia conducted by the Nanzan Institute for Religion and Culture. Earlier international workshops were held in collaboration with the Center for the Study of Science and Religion at Fu Jen Catholic University in Taiwan (October 2007) and the Sogang University Institute for the Study of Religion in South Korea (October 2008). The conference was the culmination of a five-year long dialogue between religious scholars (representing a wide range of religious traditions in East Asia) and many leading neuroscientists, cognitive scientists, medical engineers, nuclear physicists, and other science professionals, on issues such as the nature and the origin (emergence) of human consciousness, spirituality, heart, and mind. There was further development of many of the previously discussed issues that were raised at the preliminary GPSS conferences and workshops in East Asia, and this

opened new horizons for the future development of the dialogue between science, religion, and spirituality.

List of Invited Participants

Frank Budenholzer, *Fu Jen University, Taiwan*
Chan Ying-Shing, *The University of Hong Kong, China*
Funahashi Shintarō, *Kokoro Research Center, Kyoto University*
Alena Govorounova, *Nanzan Institute for Religion and Culture*
Huh Kyoon, *Ajou University Medical Center, Korea*
Iriki Atsushi, *riken, Japan*
Tom Mackenzie, *Head of Philantrust and gpss representative, Paris, France*
Sangeetha Menon, *National Institute of Advanced Studies, Bangalore, India*
William Newsome, *Stanford University, USA*
Ōtani Eichi, *Bukkyō University, Kyoto*
Paul Reasoner, *Bethel University, USA*
Curtis Rigsby, *St. Lawrence University, USA*
Satō Tetsuya, Director, *The Earth Simulator Center, Japan*
Bernard Senecal, *Sogang University, Korea*
Michael Spezio, *Scripps College, USA*
Paul Swanson, *Director, Nanzan Institute for Religion and Culture (gpss, pi)*
Tanaka Keiji, *riken, Japan*
William Waldron, *Middlebury College, USA*
Yamamoto Sukeyasu, *riken, Japan*
Yokoyama Teruo, *Nanzan University*
Yoshikawa Sakiko, *Director, Kokoro Research Center, Kyoto University*

Opening Session, Paul L. Swanson, "Welcome and Orientation"

The gpss Project in East Asia (2004–2009)

The conference began with opening comments by Paul Swanson, who gave a brief summary of the progress of the gpss project in Japan and East Asia during the years 2004 to 2009. He mentioned all major events conducted within this time frame, including international symposia and conferences that have been held in Japan, Taiwan, and Korea over the years as well as numerous domestic Japanese workshops and colloquia held at Nanzan during this period.

Swanson highlighted the fact that the gpss project was financed by the John Templeton Foundation, whose attention has shifted lately to include the "big questions," such as major philosophical problems that arise as the dialogue between science and religion takes place. For the gpss participants in East Asia,

the most crucial questions in this dialogue turned out to be the questions of the origin and the essence of human consciousness in the light of natural sciences: "What is consciousness?" "What does it mean to say 'I'?" "How did self-awareness emerge in human beings?" "What is mind, heart, spirit, *kokoro*?" and "What are the moral implications of research in brain science?"

Swanson also pointed out the linguistic peculiarities of conducting the GPSS project in non-English speaking countries. Definitions of mind, heart, and spirit vary from culture to culture, and one of the focuses of the GPSS project in Japan was on defining and working with the meaning of the Japanese word *kokoro* in relation to its possible English equivalents. Throughout the GPSS project in East Asia it became evident that when discussions on the same topics happen in different languages (in English versus in Japanese, for example), they take a totally different shape and direction.

Next, Swanson defined the main problematic of the dialogue between science and religion as the dilemma of the two conceptual extremes in the interpretation of human consciousness: absolute dualism vs. complete reductionism. Absolute dualism is problematized by the question: if spiritual and material substances exist independently, then how do they interact? If they interact, as they obviously must for us to experience them, how can they be considered two independent realities? On the opposite side of the ideological spectrum is the extreme of complete reductionism, where everything is reduced to the physical activity of the brain and body. This raises a number of ethical issues in the interpretation of human essence: are we mere machines, or are we "spiritual machines"? The biggest question here, however, is how to account for agency—an experience of an "I." Particularly, when it comes to a religious context, how do we explain the problem of free will and moral responsibility? One of the purposes of the GPSS project in East Asia was an attempt to find a way to overcome or go beyond these two epistemological extremes of absolute dualism v. complete reductionism to find a "middle path."

"Seeing is Believing?" and Other Challenges

Another major objective of the project was to examine how the human brain processes information and provides our experience of reality. In particular, the focus of many discussions has been on the unconscious activity of the brain and the formation of our behavioral choices beyond our awareness. One of the participants of the project, Tanaka Keiji of RIKEN, has frequently pointed out the fact (established through his research on the brain) that often our seemingly-conscious choices are determined by the unconscious activity of the brain. Basically, the limits of our conscious decision-making are not clear and have to be reconsidered by neuroscience (and religion).

Optical illusions provide a good example. Swanson showed an optical illusion in which certain parts of an array are perceived as being different colors, even though closer examination shows that they are "in fact" the same color. Our brains still perceive them as different colors (due to their surroundings) even after we "know" that the colors are the same. Such optical illusions clearly demonstrate that even if we are aware of the existence of the optical illusion, and even if we "know" the way something should look like, our brain interprets it differently. This makes us face the possibility of self-delusion: as humans and as scientists how much do we really know and how much are we really in control of our decisions? This provides a new perspective on the idea of "seeing is believing," and on the way we approach the issue of free will within the framework of the dialogue between science and spirituality.

Swanson also brought up some of the questions that may arise during the present conference: "What can brain science tell us about consciousness?" "What can it tell us about religion?" "How can brain science help us understand mind, heart, spirit, and *kokoro*?" "Do brain science and religion have anything to offer to each other?" "Do the findings of brain science explain the nature of your religious beliefs?" "What do your religious beliefs say to you as a scientist?" He also emphasized that instead of asking: "What can science learn from religion?" or, "What can religion learn from science?," we should rephrase the question as "What can we learn from science?" and "What can we learn from religion?" There is no need to reject the one to accept the other. Finally, where do we go from here? Are there possibilities for future collaboration? What are they?

In conclusion, all the participants were encouraged to have a fruitful debate that will help open up new prospects for the dialogue of science and religion.

William Newsome, "Neuroscience, Spirituality, and Scientific Explanation"

Following the opening remarks, William Newsome, professor at the Department of Neurobiology at Stanford University, gave an opening speech, providing a clear hands-on demonstration of how science, philosophy, and religion and spirituality inescapably merge in the scientists' labs today.

The Big Picture

The comments opened with a demonstration of a photograph of the Milky Way with a "you are here" arrow pointing at the small dot in the midst of our galaxy signifying the Earth. This was explained as a perfect illustration for "the big picture" mentality that we are currently trying to achieve: we are all here on the edge of the Milky Way, right now, where our Earth is but a dot. The world is contained within the brain, and the brain, in turn, acts on the world and changes the world. And to understand how this is done is the real project of neuroscience.

The topic of this opening lecture—"Neuroscience, Spirituality and Scientific Explanation"—highlighted the two major themes: 1. a scientific explanation of how the biology of the brain makes possible our mental lives; and 2. a meta-scientific demonstration of how traditional epistemological boundaries between natural sciences and human sciences blur today, as the dialogue between science and spirituality progresses. Many of the issues that traditionally have been the province of philosophy or theology are now coming into the laboratory, and today scientists simply cannot avoid dealing with these philosophical questions.

The Central Dogma of Neuroscience

The central dogma—or rather, the central working assumption of neuroscience—may be formulated as follows: all of our behavior and all of our mental life, including our sense of a conscious, continuing self, emerges from and is inextricably linked to the biology of the brain. This is a working assumption without which neuroscience is impossible, and professional neuroscientists will push this assumption in the laboratory as far as it can go. Why would neuroscientists believe such a thing?

"No Longer Gage": Personality Change Through Brain Trauma

A number of examples illustrate why neuroscientists subscribe to this "dogma," but one of the most famous is the example described by John Martin Harlow (1819–1907). Harlow was the first person to suggest a specific behavioral role for the orbito-frontal cortex. Harlow published two famous papers in his life. The first one was "Passage of an Iron Rod through the Head" (*Boston Medical and Surgical Journal*, 1848). The second, published twenty years, was "Recovery from the Passage of an Iron Rod through the Head" *(Bulletin of the Massachusetts Medical Society,* 1868). These dealt with the famous example of Phineas Gage, a twenty-five year old railroad worker, whose accident in September 1848 changed our understanding of the relation between mind and brain. Harlow describes the accident as follows: Phineas Gage was excavating a rock when a premature explosion drove the tamping iron—1.1 m long, 6 mm in diameter, and weighing 6 kg—through his left cheek and out of the vault of his skull with such force that it threw him on his back and fell several rods behind, "smeared with brain." Despite his injuries, Gage remained conscious and even preserved the ability to walk with assistance to the hospital. Amazingly, he survived this accident and recovered from it, but his personality was totally changed. In Harlow's papers he quotes Gage's friends and acquaintances, all of whom claimed that "he was no longer Gage. Impatient of restraint or advice when conflicts with his desires, at times pertinaciously obstinate, yet, capricious and vacillating, devising many plans for future operations which are no sooner arranged than they are abandoned in turn for others appearing more feasible." What Gage lost was the ability

to organize behavior at a high level and follow through on plans. He had an ability to make many plans but as soon as he made one plan, he would change to another plan. It was the continuity of behavior, and the continuity of purpose, and the continuity of planning that Gage lost when he lost this particular part of his frontal lobe. Harlow's analysis suggested that even those aspects of our mental life that we consider as being the highest parts of our selves are critically dependent on particular mechanisms in specific regions of the brain.

Brain Stimulation Recreates Childhood Memories

This theme has been continued in experiments over the history of neuroscience, perhaps most famously the experiments of Wilder Penfield, a neurosurgeon in Montreal who pioneered surgeries on the exposed brain of awake human patients suffering from intractable epilepsy. When searching for the "epileptic focus" so that it could be surgically removed, Penfield electrically stimulated parts of the patient's brain while the subject was completely awake. During these surgeries he made the remarkable discovery that stimulation of the temporal lobe of the cerebral cortex could cause the subject to recall highly specific, distant memories, such as their mother calling them to dinner when they were a child, or the exact music that was being played during a specific event in the subject's past. The fact that the subject had these vivid experiences when during artificial brain stimulation demonstrates that highly organized experiences are created by the electrical activity of the brain. Moreover, they can be recreated artificially by activating the appropriate neural circuits. This is why neuroscientists have come to the conviction that our highest mental and even spiritual faculties are somehow linked deeply and indivisibly to the activity of the brain.

Intersections Between Brain Science and Religion: Consciousness and First Person Reality

Newsome argued that as neuroscience progresses, there will be, and are right now, two intersections of neuroscience with religious traditions and religious understandings of what it means to be human. The first fundamental intersection between science and religion concerns the nature of the human consciousness, or, as pointed out by Swanson in his opening remarks, "What does it mean to say 'I'?" What is this sense of internal organization, internal integrity that we have; what is this conscious self-awareness? Does it play a causal role in the overall function of the brain, or as some people have argued, is consciousness epiphenomenal and all the real work of the brain is done at the neurocircuit level? Does this imply that consciousness is merely an illusion that we have about our being?

The second (related) challenge that has to be put on the table for discussion is the challenge of reductionism concerning the causal status of mind and what we sometimes call "free will." According to Newsome, the questions of consciousness and first person reality are going to be the increasingly pressing issues for neuroscientific study, and they have already come up in his own laboratory in neuroscientific experiments on non-human primates.

"Normal Neuroscience"

To illustrate how the issue of consciousness and first person reality has come up in the laboratory, Newsome showed a diagram of the way many neuroscientists look at this input-output function: from the Milky Way (the above-mentioned photograph) information comes into our sense organs and at the higher level of the cortex we actually perceive things. That is how humans can perceive a famous Kanizsa triangle[1]—even though the fact that there is no triangle present on the retina—but it is suggested by certain elements and we perceive it quite vividly. At this higher level of perception exists decision-making, and decisions that we make strongly influence movements that we plan, and then ultimately movements that we make. So, if we decide that a particular visual stimulus is a predator, we may run from the predator, but if we decide that it is prey, we may approach the object. These decision mechanisms are the key links between perception and action, and these higher cognitive functions such as memory, attention, motivation, and emotion are mediated by higher order "association" circuitry in the cortex and elsewhere.

Newsome then explained that in his lab a lot of work has been done on visual motion signals in order to explain how one perceives moving visual objects. The subjects in the lab are monkeys who are trained to do visual discrimination. There are several primary and many secondary visual areas in the brain, but there is one visual area in particular called MT or "middle temporal visual area" that seems to specialize in processing motion information. Neurophysiologists typically study cells in these areas with microelectrodes, which are very thin pieces of tungsten and platinum that are chemically etched to a tip size smaller than a human hair. The electrode is insulated to within a few tens of microns of its tip and this tip, if it gets close enough to a nerve cell, can measure the electrical action potential of this cell without interference from other cells nearby.

1. The Kanizsa triangle is named for the Italian psychologist Gaetano Kanizsa, who published his findings of the optical illusion in a 1955 edition of the Italian journal *Rivista di Psicologia*. He noticed that we see a white triangle on top of and partly occluding disks and another triangle. The triangle, however, has no physically measurable existence although they appear to observers as significantly brighter than the background. [A.G.]

Neuroscientists put these microelectrodes in the brain and record the activity of cells and notice that the cells in MT seem to code the direction of moving objects. What that means is that a visual stimulus like a bar of light, moved across the retina in one direction, will illicit a burst of electrical activity. But the same stimulus moved back across the retina in the opposite direction elicits no electrical activity. So, this cell in the cortex is extracting information about the direction of motion by virtue of the circuitry between the retina and the cortex, and it is broadcasting that information to the rest of the brain.

To demonstrate how this works, Newsome showed a movie depicting a circular aperture with random dots moving in the aperture in eight different directions, one after the other. In a movie, these random dot movements are accompanied by a sound track, where every sound represents an action potential from a single neuron. The movie clearly depicted the direction selectivity of the cell: motion down to the left (the "preferred direction") elicited a vigorous burst of electrical activity. As the direction of motion departed from the preferred direction, the cell's electrical activity became weaker and eventually disappeared altogether for motion up and to the right. Newsome clarified that the cell was recorded in his laboratory while the monkey was actually looking at the screen with a small electrode in this MT area in his brain.[2]

Newsome then described a 2-deoxyglucose imaging experiment that was done by Rodger Tootell and Richard Born. The experiment showed that direction selective cells in MT are organized in spatially segregated clusters, or "columns": groups of cells selective for upward motion were separated from groups of cells selective for downward motion by distances on the order of a few hundred microns. Neuroscientists have known about these columns since 1984, but what they did not know is whether the information contained in these columns is actually used when the monkey makes judgments about motion direction.

Newsome and his team trained monkeys to do a direction discrimination task. The experiment can be summarized as follows: the monkey views a display of moving random dots on a TV monitor, and his task is to determine whether the flow of dots is in one direction or its opposite (e.g. up vs. down). At the end of the trial, the monkey reports the perceived direction of motion by moving his eyes to one of two visual targets corresponding to the two directions of motion. For example, if he felt he saw the upward motion, he moves his eyes to a target positioned above the dot display. Basically, the monkey is simply flicking his eyes to one or the other target to tell the scientists the direction of motion that he

2. Newsome emphasized that the brain has no pain receptors, so these electrodes do not hurt either humans or animals. That is why neuroscientists can record this neuroelectrical activity in the brain without causing any physical pain or discomfort to experimental subjects. The monkey simply sits there, looks at the screen and allows neuroscientists to study the cells.

sees while the computer measures his eye position. The computer always knows where the monkey's eyes are, and if the monkey makes the choice correctly, then he is rewarded a few drops of juice that he likes. If he chooses incorrectly, he does not get any reward (which he does not like). Therefore, the monkey is highly motivated to make the correct choice. The scientists put microelectrodes into the middle of one of these columns: red means down, green means up—this is determined beforehand, before the monkey begins to discriminate. Then the scientists ask the monkey to look at dots and tell them, did he see up, or did he see down?

The scientists conducted two sets of experiments: first, they allowed the monkey to report the motion of the dots when the electrode was merely resting in the targeted column and not affecting the cells; second, they electrically stimulated the cells while the monkey viewed the random dot stimulus in order to input an artificial signal into the down column or into the up column—to see if this changes the monkey's decisions. The results show that electrical stimulation changes the monkey's decisions dramatically. In a situation where, for example, the real motion is upward, but scientists stimulate a "downward" column in the monkey's brain, they can make the monkey say "down" and they can get it to happen very reliably.

To summarize, neuroscientists can override the subject's decisions by using the electrical stimulus in these very fine-grained circuits, which are only about a hundred or two hundred microns in diameter. As a result, while the monkey is looking at the dots and is trying to decide if the dot is up or down, neuroscientists, by stimulating an MT, can predict or can cause him to say "up" or cause him to say "down."

Newsome clarified that all that he has described so far could be categorized as "normal neuroscience." The experimental results described above were astounding when neuroscientists first discovered them—because until then no one believed that these tiny little circuits in the brain could be responsible for a complex perception like the direction of motion in dots—yet, it turned out to be true. However, so far it is "normal neuroscience," where neuroscientists manipulate circuits and signals in circuits and question the effects of this artificial manipulation on behavior.

Beyond "Normal Neuroscience"

Newsome argued that when it comes to these kinds of experiments, neuroscientists must ask a question that goes beyond "normal neuroscience." The question that arises here is "What does the monkey *see* when an MT is stimulated?" We can record his choices and we can reward him or not reward him according to his choices, but we do not know *what* he actually experiences. Maybe, if the monkey could talk to us, he might say, "Hey, I saw upward motion on that trial

and I reported upward motion correctly, why didn't I receive my reward?" Or the monkey might say, "I saw downward motion but I reported upward motion, I do not know why I did that." The above are two very different conscious subjective experiences that in the end yield the same data that we record on our computer.

Currently, we do not know how to answer this question with experiments on monkeys. Maybe we can answer this with experiments on humans, Newsome suggested, adding that he personally wants to have his MT stimulated. This, however, would induce all kinds of ethical problems and we can debate this issue, but it is a slippery slope.

However, curiously enough, there are some neuroscientists who will say, "Is this question (of what the subject really sees) even an important question?" Newsome argued that it may be the most important question for future neuroscience. We can well imagine the future—forty years from now, or a hundred years from now, or two hundred years from now—when we can understand all neural events that code dots on the retina, and how all of these signals are processed in the cortex; we know exactly from a mechanistic point of view how decisions get made, and how the decision informs and creates an eye movement up or down. However, what if we know all of that, and we still do not know this: what does the monkey see when his MT is stimulated? Maybe we can explain the entire matter in terms of physics from input to output, but we still do not know what the subjective correlate is.

Newsome insisted that his question for all of us today is: how would we feel, how would we think about that? Indeed, some scientists would say, "If you know everything from input to output, you have solved the problem. You quit. You go home. You go to the next problem." But some of us would say, "I do not know [yet], this is an important question." What we really want to know is how *subjective mental life* arises out of the brain. And if we cannot answer this question, then we have missed something important.

Newsome strongly emphasized that the operation of the brain and the things we want to know about the brain are more than simply the physics of the neural interactions—it is about how the mental life that we experience is related to that physiology. This is a crucially important question and, of course, related to what David Chalmers would call "the hard problems of consciousness." Newsome clarified that his purpose in bringing this up at the conference was that he himself does not have an answer to that question. Moreover, his purpose in bringing it up for discussion was to demonstrate that this kind of question is not merely a philosophical abstraction. These are the questions that hit neuroscientists in the face increasingly frequently in the lab. And this issue is becoming a matter of urgency for neuroscientists to try to answer. As Thomas Nagel beautifully sums it up in his essay, "What Is It Like to Be a Bat?":

> If the subjective character of experience is fully comprehensible only from one point of view, then any shift to greater objectivity—that is less attached to a specific viewpoint—does not take us nearer to the real nature of the phenomenon; it takes us farther away from it.
> (*The Philosophical Review* 83/4, October 1974: 435–50)

And this is a real puzzle for neuroscience because we are traditionally interested in objective third person knowledge. And, as it had been previously discussed at the prior GPSS conference in South Korea in 2008, this cleavage between objective and subjective is probably the source of some of our errors, and there may never be any purely objective knowledge or purely subjective knowledge but there may always be a blend of any kind of knowledge that we have. This raises the issue that there may be fundamental limits on what we can understand about conscious first person experience using only the third person methods of contemporary science. Therefore, we simply have to think hard about that.

Perhaps, the deepest question of all would be: why should conscious experience result at all from the electrical activity of nerve cells? How can matter become conscious? How can collections of nerve cells become conscious? How can that happen? Newsome himself admitted having "no earthly idea how that happens."

On the Challenge of Reductionism: The Causal Status of Mind and Free Will

The second challenge that Newsome brought up for discussion was the challenge of reductionism, namely, the causal status of mind and what we sometimes call "free will." He specified that this issue has also come up in his lab because of the studies in decision-making. Newsome and his team study motion perception, but over the last ten or fifteen years they have done a lot of work on the neural mechanisms underlying decision-making in monkeys. As demonstrated above, even the simple little motion task involves decisions. Neuroscientists think of MT and the motion cells in MT as recording evidence about what is out there in the world. But the monkey has to make a choice. Is a particular set of dots moving up, or down? The monkey has to make a binary choice—almost like a jury in a legal proceeding. Evidence gets presented but the jury has to decide "guilty" or "innocent." And that decision-making level is different from the sensory evidence level. Therefore, Newsome and his team have studied decision-making using this perceptual task and they have discovered signals related to the monkey's decision in the lateral intraparietal area (LIP) of the parietal lobe and two areas in the frontal lobe, as well as in the structure in the midbrain called superior colliculus.

Newsome stated that he and his team believe that these brain areas act together in a network to make or to form these kinds of decisions. More recently, he and his team have been studying value-based decisions. Value-based decisions should be understood here in economic terms: they refer to a "reward" or "acquiring resources." In the case of a monkey, it would like to acquire juice resources, so when the experiment starts at the beginning of the day, the monkeys are thirsty and they work for this reward until they are no longer thirsty. Finally, the monkey will have enough juice and quit: unfortunately, he does not care about science, nor does he care about truth. He only cares about the reward, and the value is in the juice.

In Newsome's value-based decision-making experiments, the monkey starts off with his eyes in the center of the screen and then he has to choose the green target or the red target. It is important to remember here that there are no dots to tell him the correct answer; in fact, there is no "correct answer." This is what is called a "free" choice task with the word "free" in quotes here because it must be clearly defined first what "free" means. Operationally, it means that the monkey can choose whichever one he wants and the scientists reward the monkey probabilistically. Thus, they may reward the monkey with two thirds probability on the green target and one third probability on the red target. And the monkey has to figure this out by trial and error. When the monkey figures this out he gradually comes to match his choices or his choice rate to the reward rate. As long as the scientists reward him two thirds of the time, the monkey makes two thirds of his choices of the green and one third of his choices of the red.

If the scientists "change the world" and reverse these contingencies, the monkey will sense this and he will come around. Basically, some part of the monkey's brain is doing some reasonably sophisticated math, estimating the current probability of receiving a reward on each target. This does not mean that the monkey is writing equations, but still in some part of his brain he is estimating probabilities very accurately. This effect was first discovered by Richard Hernnstein while working with pigeons. Hernnstein called this the "matching law": the animals match their choices to the likelihood of getting rewards.

It turns out that neuroscientists can build nice quantitative models of the animals' decision-making process: how the animal estimates probability, how the estimates of probability get transformed into probability of a choice, and how final choices get made at the end. For example, in a simple two-choice task, such a quantitative model can predict the animals' choices with about 82-83 percent accuracy. There is some unexplained variance there but this is still considered a high accuracy. But the question about free will still arises here. Neuroscientists can build the models that predict decisions this way, and if they can begin to understand "Free" decisions in terms of neural signals in particular brain areas—what does it say about free will?

As one wag said, "I am not a fatalist but even if I were what can I do about it?" Neuroscientists believe that there are mechanisms that underlie intentionality and choices and they want to uncover these mechanisms and elucidate them. But as people, as we start thinking about the highest levels of human being, about human behavior, and what is meaningful about it, we tend to be skeptical about bottom-up determinism—that all of our choices are completely determined by mere bottom-up actions of neurotransmitters and action-potentials, and that we are sort of walking machines and we have very little choice in the matter—that is not something that seems intuitively appropriate.

Quantum Mechanics and Free Will

According to Newsome, some scientists have tried to leave room for free will by resorting to quantum mechanics and the quantum brain. He insisted that, in his opinion, this is not a good solution to the free will dilemma. It is not a good solution for two reasons.

First, according to the physicists and biophysicists that Newsome has consulted, the macromolecules that generate electrical currents in nerve cells are too big for quantum effects. Therefore, quantum mechanical opening and closing of channels does not seem to be a realistic way for quantum effects to get into brain function. Another reason why quantum effects are not an attractive solution to the free will dilemma is that they are randomly probabilistic. In the words of Newsome, "Randomness is no more intrinsically meaningful to me than complete determinism, and I actually don't want random events occurring in my brain when I cross the street out there: I do not want to see the car with 99 percent probability and miss it with one percent probability. I want it 100 percent and I want a completely reliable mechanism. This is why I am not very attracted to quantum accounts of decision making."

Newsome insisted that we have to rethink our definitions (understandings) of free will, intentionality, and choice. Apparently, some people put a very stern definition of freedom as being "uncaused" in a sense that if there are any causes to your behavior then you are not free. As Newsome put it, however, "I do not believe in freedom in a sense of being uncaused. In all of our behaviors there are underlying causes and I do not think that magic happens in the brain. Perhaps, freedom is a bad word to use—but what we really want is some sense of *self-determination*: we do not want complete bottom-up determinism, and we do not want randomness popping up unrelated to our goals, desires and personal history. What we want is some sense that this organism that we are, that whole that comprises the human person, has self-determination or some sense of autonomy. And the question is whether this sense of self-determination or autonomy is consistent with the physical understanding of the brain.

And I think that it may be and the secret may lie in this concept of emergence in complex systems. I am aware of the range of problems connected with the concept of emergence but I believe that, if properly understood, it may offer some helpful ways to think about this."

Newsome's main argument was that the standard reductionistic paradigm in science is insufficient for understanding high-level phenomena. Within this paradigm, scientists tend to take things apart into smaller and smaller pieces, accompanied by the belief that when we successfully take a high-level phenomenon apart into small scale mechanisms, that high-level phenomenon is no longer relevant or no longer has power because the more fundamental, more truthful, level of understanding is the low-level mechanism. This new low-level mechanism, in turn, becomes the high-level mechanism to be taken apart into yet lower level mechanisms, resulting ultimately in a complete reduction to fundamental physics (quantum mechanics). He argued that this paradigm for understanding human persons is impoverished—not wrong, but impoverished. In principle, if we were smart enough and if we had enough information we could write Schrödinger wave equations that would predict the motion of every atom in this conference room for the next twenty minutes—probabilistically. However, the Schrödinger wave equation will not know anything about the people in the room. It would not know anything about their nations of origin; it would not know anything about the questions that brought the participants of this conference together; the Schrödinger wave equation will not know anything about the feelings of jet lag that they have or interests that they have. The wave equation can predict the motions of atoms but the wave equation is impoverished when it comes to the high-level phenomena. Therefore, the emphasis on reduction and throwing away higher levels is where the problem exists, according to Newsome. Perhaps, this poverty is our real problem in approaching not only the free will phenomenon but also in thinking about consciousness and maybe about some other emergent phenomena.

To illustrate this, Newsome gave one example, which comes from the world of neural networks. Neural networks were known back in the late seventies and it was the mid-eighties when they really took off. Neural networks are little computing circuits that are implemented typically in digital computers. The key idea about the neural network is this: it has a series of units, little computing units, and each unit adds or subtracts its inputs to produce an output. These units are typically organized in three layers—input, hidden layer, and output. The connections between units in the successive layers are initially random, but are adjusted during a "learning" process according to a specific "learning rule." After hundreds or even thousands of learning trials, in which the connections between units are continuously modified, the network converges on the solu-

tion to the problem at hand, whether it be converting written English to spoken English, or identifying faces.

For those not in this field it is hard to imagine the sense of amazement that accompanied this development in the 1980s. Until the 1980s our standard approaches to artificial intelligence were programming things from first principles, trying to write programs that would take visual images and interpret it as a camera or a projector or a bottle of water. In fact, in the 1960s people were very optimistic about that: they said within twenty years we will have programs that see as well as a dog or will be as smart as human infant, which, of course, was terribly wrong. These things are really hard to do from first principles. However, in the 1980s when these neural networks came into being, they could solve problems that have defied the very best human programs for twenty years. So, these are extremely powerful networks and the secret is in this learning; the secret is in having all these connections, and these connections can be modified depending on how close the output is to the correct answer. Neural networks, thus, were used in speech recognition, and they were used in artificial vision. They were used and are still used as a means to solve very many practical problems.

Newsome recalled standing in front of a poster of a neuroscience meeting back in the mid-1980s with someone showing him one of these new discoveries, specifically that this neural network could predict the outputs of nerve from the inferotemporal cortex—a high level visual area of the cerebral cortex. It was clear that neural networks could do these predictions—it was obvious from the data. Newsome's question, however, was "How does the network do that? What is the answer? What is the answer to the problem?" Unfortunately, all they can tell you is that the answer is in the weights of connections between the computing units in the network. Newsome then showed a diagram depicting these units and their weights: this is hidden unit one and the diagram shows the weights of hidden unit with all of its inputs; the color black indicates a negative weight, white means there is a positive weight, and the sizes of the circles show how strong the connections are. And this is the final state of the network after the problem is solved: hidden unit one has this pattern of weights, hidden unit two has this pattern of weights, and so on. In a sense, this really is an explanation because we can take this information out and move it from a computer in America to a computer in India, or a computer in France or wherever, and we can get the exact same solution and we can really solve the problem. In other words, we understand everything about this in a physics sense—in a physics sense we know exactly what these weights are, we can reproduce them; we can ship them to any other scientist in the world; we can imbed them inside of Microsoft Word or Adobe Photoshop (and there are real neural networks imbedded into software inside of the Adobe Photoshop), and so on. In summary, in a physics sense this is totally understood.

Nevertheless—and this is very important—the network is deeply mysterious in another sense because when we say, "this network can receive written English and produce spoken English words correctly … that network has solved a very important problem!" and we say "how does it do that?" Then someone shows us those patterns of weights and we say to ourselves "that's not the kind of answer I am looking for!" The conviction is that there is a deeper level of understanding we can get to—an understanding of principles rather than just connectivity diagrams.

Is There an Answer?

Is there an answer? Or maybe there is no answer? What we really want to know is the principle involved—the computational principle by which the input is transformed to the correct output, in other words, even if the network is completely understood physics-wise, it is not understood computationally-wise. Some have called these networks, tongue-in-cheek, "Know-nothing networks" because in the end even if the network has solved the problem we still do not know the answer to the problem. The network somehow knows but we do not know. Therefore, we can say that there is a gap between the physics-level understanding and the kind of understanding that we really want. And that gap is a signature of an important emergent phenomenon or an important emergent level of understanding—it is essential to understand.

There is a very important point here that cannot be overemphasized: Talk of emergence—in this context at least—does not involve anything "spooky." There is no magic, no dualism, no mysterian doctrines, no ghost in the machine—the network is totally understood in terms of physics. But there still is a gap between the physics-level understanding and the higher-level understanding that we really want to achieve. In fact, this kind of gap is more common than we realize and exists throughout nature. In the case of cells, for example, we understand a lot about how cells work in terms of genetic codes, proteins, enzymes, chemical reactions, signal transduction, motility, metabolism—all of these things we understand about cells, but if we try to reduce the life of single cell organisms totally to this level of understanding, we would miss higher level phenomena such as predator-prey relationships, foraging, and symbiosis; how would we even describe the concept of a predator in molecular terms? This is a higher level of relationship that is important to understanding nature but it ultimately defies reduction.

The same is true for computers. We can understand a computer in terms of its circuits, its transistors, its capacitors, its resisters but, arguably, the most important level of understanding of the computer would be the computing strategies embedded in the software—for-loops, if-statements, subroutines, etc.

These all are high level phenomena, but they exert causality in a very real sense: the software exerts causal control over individual components, and while the software is implemented in the computer hardware, we cannot totally reduce computing concepts to transistors and capacitors without losing something important. Again, there is no magic, no ghost living in the computer that does the "smart stuff"; the argument here is about the appropriate level of description of the system and where the real work is getting done. In this sense, the neural network example discussed above is particularly powerful because we do not know about the high-level, end-state of the network in advance. We know the learning rules, but the network ultimately achieves a solution that we cannot predict in advance (if we could, there would be no need for the network!). In the case of cells, however, we do know these high-level relationships in advance. We knew about predator-prey relationship, symbiosis, and parasitism long before we knew molecular biology. Thus, in biology, we sometimes fail to perceive the sharp limitations of reductionist analysis because we take the higher level phenomenon for granted—they are the things we already know that we are trying to *explain*. In contrast, the neural network example, in which we can understand the mechanism completely but still not understand the logic of the network's solution, forces us to grapple with the limitations of a purely reductionist analysis.

As Newsome put it, this kind of gap—emergent gap—at the very least may be defined as epistemological. To understand something we have to bridge that gap, it is not enough to reduce it—we have to understand a higher level as well. However, there is another deep question as to whether it is ontological, whether the higher levels of organization are real in an ontological sense, and in what sense we can regard them as doing causal work. Applying this logic to the brain, we may come to regard consciousness as a higher level state of a neural system. Consciousness may be real in the same sense that software is real. And at this higher level of configuration, the key question for us to understand is that of "downward causality"—how higher level configurations exert causal influence onto the lower levels. If a higher level system configuration exerts causal downward influence into the lower levels, then maybe that is what we really mean by human autonomy and self-determination. If so we should eventually get rid of the term "free will" and make the key word "autonomy."

Is This Science or Is This Philosophy?

Newsome acknowledged another question to struggle with: is this science we are talking about here or is this philosophy? Does this have real implications for how we do science and how we understand science or is this something other than science? In Newsome's opinion, there are some indications that this is science; moreover, understanding the nature of human freedom is the single most

important problem facing the neural and behavioral sciences: what is a human really, what is autonomous behavior? And how to interpret biology, understand biology fully in a way that gives appropriate weight to mechanism and also gives appropriate weight to the higher level, to the behavior integrating the whole system, including the system that is a human being? Obviously, this is important for reasons of human dignity, and this impacts our legal theories about responsibility. However, what about the hypothesis that this is important for science?

Newsome closed his talk with a quote from J. B. S. Haldane, a famous evolutionary biologist of the mid-twentieth century: "If my mental processes are determined wholly by the motions of atoms in my brain, I have no reason to suppose that my beliefs are true … and hence I have no reason for supposing that my brain is composed of atoms." There is a circular problem that we have, emphasized Newsome. "If I believe completely in bottom-up determinism, then everything I am thinking right now is simply the product of bottom-up motions of atoms in the brain. But what does it do to the concept of truth? How is anything true then why would I believe that my brain is composed of atoms?" Science itself demands that humans have a higher level capacity to evaluate evidence and make rational judgments about what is, and is not, true about the world.

Paradoxically, then, understanding human freedom, or autonomy, is just as important for science as it is for the legal system or for ethics. According to Newsome, it is impossible to do science and to talk about rational approaches to truth—moreover, it is impossible even to define "truth" unless there is some way to have independent judgment. These are philosophical questions, these are questions of religious concern, but these are questions that are emerging right from the laboratory. We can return to that question about consciousness: what would the monkey say if he could talk to a scientist stimulating his brain? That is a question coming out of the laboratory. And we can return to our scenario of an imagined neuroscience 200 years into the future in which neuroscientists can predict decisions with one hundred per cent accuracy and still not know exactly what the monkey *sees during brain stimulation*—these are the questions that are coming out of the laboratory. They are real examples of issues where deep discussions between scientists, humanists, and adherents of religious traditions are critically important. This is how we can arrive at a realistic, accurate, scientific understanding of nature and the human person, and discuss the essence of human dignity and spirituality.

Discussion

The discussion began with a question by Paul Swanson, referring to the "central dogma of neuroscience" presented by William Newsome as follows: "all of our behavior and all of our mental life including our sense of a conscious, continuing self emerges from and is inextricably linked to the biology of the brain."

Swanson asked Newsome to clarify "how far the brain goes," specifically, does it stop at the base of the head, or does it extend through the neural network that goes to other parts of the body? How do we technically delimitate the brain? And the extension of this question is: how about the "gut feeling"? Some people refer to the "gut" as a second brain because there are so many neural networks there, perhaps even more than the brain in our head. But can we really say that we "think" with our gut?

Newsome responded that, to his knowledge, indeed there are more neurons in the enteric nervous system than in the central nervous system. The enteric nervous system has a series of ganglia down in our gut that monitors internal process and controls many vital functions such as peristalsis of our gut. Those are real parts of the nervous system and they provide important information back to the brain about the status of the internal organs. The enteric neural system sends signals about the concentration of blood metabolites that control hunger and thirst, and when we feel pain in our gut, it comes from the enteric neural system. Newsome stated that the enteric system is a vital part of our nervous system because it is the key source of our awareness about our internal life status. However, the next question arises here: what about the body itself, do we simply restrict this awareness to the nervous system or do we extend it to the body itself?

Newsome specified, however, that it is probably not scientifically correct to say that we "think" with our gut or "make decisions" with our gut but, perhaps, our gut influences the decisions that we make. It provides important inputs to the decision-making process and, in fact, there are different levels of decisions. We touch a hot burner and we withdraw our hand and that information did not even go to the cortex—it only went to the spinal cord. So, certain very simple choices and actions are processed in the spinal cord and there are higher levels of decisions in the sub-cortical regions of the brain. But as for intelligent decisions of the highest level: whether to go to college or not, whether to accept a new job—probably those are processed at the cerebral cortex, at the highest level. Although, we still tend to use this expression in English, "Go with your gut" when we think of important life decisions such as getting married or going to graduate school.

One of the participants further commented that in Japanese martial arts sometimes we are taught to "think from your stomach" when reacting to an attack, so this "gut feeling" concept may be cross-cultural.

What is the Neural Basis for Intuition?

A new question naturally followed: where does intuition fit into the scheme of things here? Newsome argued that intuition is very contextual. Intuition comes from a summary of life experiences. Our intuitions about right or wrong, or whether we are likely to get more meaning from this choice or that choice—it is

a summary of life, it is shaped from our education, from our family background, from the culture that we lived in, it is shaped from the professional training that we have; it is shaped from every day experiences of positive or negative consequences of our choices—and that builds up these contextual states of the brain that make us more inclined to certain kinds of behavior than others. In sum, Newsome concluded that intuition is going to be almost at the very highest level of decision-making as the summary of a life-long series of inputs that expose us to one kind of behavior rather than the other.

One of the participants further commented on the difference between inspiration and intuition, stating that intuition occurs rather frequently but inspiration occurs rarely and there is a biological difference between the two. Thus, according to studies in intuitive generation of the best next move in board game players, inspiration occurs in the basal ganglia, which is thought to be the center for habit. So, for professional top class players, generation of the best next move in a given situation is a kind of habit. And, perhaps, in the case of professional world-class mathematicians (or any other top intellectuals or professionals) inspiration resides in the basal ganglia.

The opening session ended with the final comments by Sangeetha Menon, who emphasized once again the importance of the issues discussed by William Newsome. She referred to the concept of "non-substantiality of self," popular among contemporary neuroscientists as highly problematic and basically invalid for understanding human nature and complex behaviors. Indeed, if the concepts themselves (including scientific concepts) are devised by our selves, how can any of the concepts we use be really true? In conclusion, she expressed hope that even if we cannot find a straightforward answer at this point, these are the most exciting questions that will help the present discussion to go forward.

Session I: Chan Ying-Shing, "Spatial Navigation and Perception"

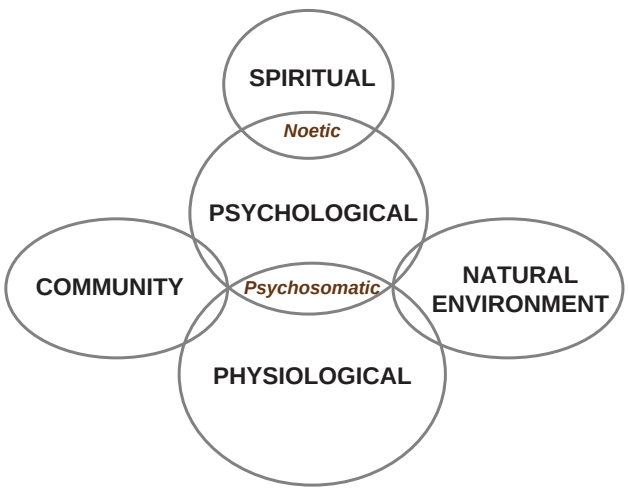
The first session opened with a presentation by Professor Chan Ying-Shing of The University of Hong Kong, China. In his talk, he described the studies in biological mechanisms of postnatal development of special navigation in the brain based on the neuroscientific experiments on postnatal rats. While rich in detailed technical explanations, Chan's talk was aimed at a high-level understanding of human self, with fundamental questions in mind, such as "What is in the brain? What is spirit?"

As a practicing Protestant Christian, Chan challenged the audience to revisit the Biblical account of creation in terms of neuroscience. God creates man with a body, mind, and a spirit. But then, what is our faith and what is the correlation between self, community, and family in our daily life? What factors and what environmental impulses affect the human body and human mind (up to the level of

the spirit) and influence our sense of inner balance? As neuroscientists we rarely think about such questions, but it still affects us in daily life. With these questions in mind, Chan framed his presentation so as to connect the realm of research in spatial navigation with the religious concept of spiritual orientation in life.

According to Chan, to achieve sensory-motor coordination in spatial navigation, a map of spatial information in the brain is required. The adult pattern of spatial information emerges during postnatal development. This involves the progressive development of neural networks for recognition of spatial orientation and differential processing of sensory cues of balance with postnatal expansion of experience. In the embryonic stage, the connectivity of hindbrain neurons that are related to the sense of balance is controlled by molecules in the surrounding matrix. These molecules pace and confine the projection patterns of these cell processes and thereby the connectivity of the neural network.

Postnatal rats recognize different horizontal orientations at different time points during maturation. Such a temporal difference in postnatal recognition of horizontal orientations is corroborated by the responses of central vestibular cells. These cells show changes in their postnatal capacity in deciphering spatio-temporal cues for the estimation of head orientations. Central vestibular cells in neonates when compared with cells in adults have a high resolution in spatial coding. The maturation of sensory transmission in central vestibular cells is also modulated by 1. connectivity with the cerebellum, and 2. crossed neural connection between two sides of the hindbrain. Altogether, spatial and temporal integration of sensory cues by central mechanisms is crucial for coherent recognition of spatial orientation. Abnormal locomotion and circling behavior therefore reflect consequences of 1. injury of one's inner ear, and 2. mutations that lead to deficiency in vestibular sensory nerves at an early postnatal stage.



Central vestibular cells of postnatal animals detect rotational movements earlier than translational movements. A cascade of maturation time for the recognition of three-dimensional spatial orientations occurs along the central vestibular pathway. The relay stations include the vestibular nucleus (the first relay from the inner ear to the brainstem), cerebellum (for learned rote movements) and thalamus (a clearinghouse for the cerebrum and hippocampus that support cognition and memory). Internal spatial maps of three-dimensional orientations have been revealed in higher relay centers.

In the normal brain, excitatory chemicals between brain cells have been implicated in playing important roles in functions ranging from learning and memory to neural network building in development. Sensory inputs from the inner ear activate specific receptors on the membrane of central vestibular cells. Developmental acquisition of specific receptor subtypes in these cells is crucial for spatial behavior. Apart from these excitatory inputs, an inhibitory chemical messenger also plays an important role in the development of space recognition in the balance system. The activity of a central vestibular network is therefore shaped by the combined actions of excitatory and inhibitory synapses.

The establishment of internal spatial reference in mature animals is deterred with the blockade of excitatory or inhibitory transmission in the vestibular circuitry during a neonatal period of susceptibility. These animals also show deficits in motor coordination and learning. These suggest that the brain refines its network for spatial recognition and expression of spatial behavior within a time window of postnatal development when the brain must obtain critical experience to develop properly.

In adult animals, a deficit in spatial navigation was evidenced with lesion of a specific thalamic region (a relay station in the central vestibular pathway), indicating that vestibular information is essential for spatial navigation. A similar navigation deficit was also reported with the lesion of the hippocampus, implicating the importance of working memory in spatial navigation. Since the thalamus is known to send projections to the hippocampus and the forebrain, the integration of orientation cues and learning-induced modification of transmission efficacy should play cardinal roles in spatial recognition and spatial navigation. Further investigation is needed to unveil how such learning and cognitive processes are transformed into psychosomatic domains for navigation in humans.

In the concluding part of his talk, Chan took up the problem of spatial navigation on the entirely new level of "spiritual navigation." He challenged the participants to think about space and time in non-reductionist terms and to rephrase the question of "What are we as humans?" in terms of direction in life that we seek. Perhaps, he suggested, we can achieve the solution to this problem with the help of neuroscience? Chan referred to the works of Andrew Newberg

and his colleagues who did comparative neuroscientific research of Christian and Buddhist meditation in order to discover neural pathways in the brain that correspond to these experiences. When conducting experiments on Christian and Buddhist meditative subjects, Newberg discovered increased neural activity in the prefrontal cortex of the right hemisphere of the brain and decreased activity in the inferior parietal lobule, which is considered to be the orientation association area of the human brain.

Chan suggested that, since neuroscientists discovered that the sense of vestibular balance is in the inferior parietal lobe of the prefrontal cortex and can be altered by meditation [hence, the experience of being "one with the universe" or "one with the Absolute"], perhaps we can say that the right part of our brain is linked to spiritual phenomena. The challenge is to link these studies to neuroscientific research in space navigation. He argued that the main question that we need to ask ourselves (as neuroscientists and as believers) is: what is it inside our body, or inside our brain, which causes us to search for spiritual direction in life? How can we connect our mind (brain) to our heart?

Huh Kyoon, "What is the Brain? Searching for the Human Mind"

The second presentation of the first session was by Dr. Huh Kyoon from Ajou University Medical Center in South Korea. In his presentation, Huh argued that owing to spectacular advances in neuroscience and neurotechnology, the brain is a mysterious black box no more. The brain is no more than an adaptive biological organ produced by the evolutionary history of the planet Earth; is no more than an assembly of a variety of nerve cells (neurons) which basically perform information processing as the unitary component. What remains to be elucidated is the complexity of the large-scale structure consisting of countless neurons and cognitive functional modularity with rich interconnections.

Using a variety of illustrations, Huh listed and explained some of the recent advances in neurosciences:

1. Linking anatomy, physiology, and psychology utilizing detailed imaging technologies of the living brain
2. Revealing the parts of the genome and distinctive DNA sequences and their unique evolutionary functions unique to Homo Sapiens
3. Regenerative medicine to restore the structure and function caused by aging and disease processes using molecular neurobiology and genetic engineering
4. Smart drug development aiming for precise sub-cellular targets and individuals not only to treat the brain disorders but also to improve the mental state and happiness of a normal person

5. Huge progress in human-machine interfaces such as implanted neuromorphic chips, neural prostheses, and brain stimulators to extend human capacity
6. High-powered computer technologies that enable us to do simulation and manipulation of brain circuitry
7. A whole new field of "Social Neuroscience" that studies the brain events involved in recognizing and responding to the feelings and emotions of others.

Many of these developments are raising unprecedented ethical, philosophical, and spiritual concerns, such as the issues of mind-reading through fMRI, pharmacological enhancement of intelligence and emotional status, prediction of human behavior, and mind control. However, there is still a lack of integration across the various levels of analysis.

When compared to other scientific disciplines, neuroscience directly addresses the daily lives of human experiences including intelligence, cognition, emotions, mind, and even religion and spirituality. It also has a strong tendency to affect public value systems, popular culture, and public policy in regards to the planning of human affairs. However, the influence of contemporary neuroscience on society is tightly bound with the strict naturalization of the human mind, based on its methodological reductionism, ontological materialism, and neural determinism.

Huh argued that the idea of "mind uploading" is a speculative but straightforward logical endpoint of contemporary neuroscience and neurotechnology. Many neuroscientists, philosophers, and even popular public opinion share the notion that the mind arises from the activity of information processing in the brain and that this can be dissociated from the biological body, just like the analogy of the software and hardware of a computer. Mind uploading (or Whole Brain Emulation) is the hypothetical process of the scanning and mapping of the human brain, and transferring its functional state into a different computational device. Once uploaded, the computer system runs a simulation model so well that the same intelligence, memory, personality, identity, and experiences of the original brain are to be maintained.

Once mind uploading is successful, the next step would be the creation of an "artificial body" that carries a computer with an uploaded mind just as our bodies carry our brains. Or, the other alternative is the creation of "artificial reality" (or Whole World Emulation)—to make an environment in the computer for the uploaded mind to live in. Equipped with a proper artificial body and artificial reality, an uploaded mind would provide incomparable advantages to humanity in terms of flexibility, freedom, a high level of pleasure and sense of well-being,

as well as an indefinite life span unobtainable in our current evolutionary biological system.

In fact, the concept of mind uploading is quite extravagant due to its extremely speculative nature and futuristic time frame. But the availability of high resolution brain scanning technology and artificial neural network systems with increasing amounts of memory storage and computational power is growing with exponential speed, and some claim that mind uploading will become a feasible technology around 2050. Outside the realms of science and engineering, the idea of mind uploading generates many ground-shaking challenges that demand exhaustive thought experiments by social, ethical, philosophical, and religious communities.

Among those challenges are:

- What is it like to have an uploaded mind?
- What is personhood and identity?
- What is freedom and privacy?
- Do we need gender, marriage, and family in artificial reality?
- Do we need any further biological evolution?
- Is life going to be a two-stage operation (biological and uploaded)?
- Will there be any laws, either physical or legal?
- What is life?
- What is death?
- What is a person: post-human identity?
- Will there be suffering in the uploaded mind?
- Is it a non-religious path to salvation and enlightenment?
- Will "religion" survive?

In conclusion, Huh stated that the question of "What is the brain?" goes far beyond natural sciences. It is a social issue as well as an ethical issue. It is often said that we live in "the century of the mind" and "the neuro-centric age," where the main question is "What is the brain?" The answer to this question will eventually lead us to shape the future of the human mind.

Discussion

Questioning the Role of the Hippocampus and Thalamus in Emotional Navigation

The discussion opened with a question by Sangeetha Menon to Chan Ying-Shing regarding the hippocampus and thalamus as deciding brain components

in special navigation. She summarized Chan's presentation as focussing on the concept of navigating or moving towards a goal (in the case of rats, towards a reward such as food). In addressing the arguments for spiritual orientation presented by Chan, Menon pointed out that there is a place for the hippocampus and thalamus in deciding the *emotional features* of the brain. In this respect, she enquired, when we talk about "navigating towards God," would that mean that there is a role for emotions to play in navigating towards better states of your own existence? Maybe there is a place for the hippocampus and thalamus in emotional navigation—not just physical special navigation—to help us reach better spaces of our own consciousness?

In response, Chan explained that, indeed, from a neuroscientific point of view, there is a specific network or neural circuitry responsible for spatial orientation in animal brains and perhaps this concept may be extended to human brains. Neuroscientists normally focus on particular components of the brain, and, in fact, the brain is highly compartmentalized. As a result, there is still a gap between some of the facts that neuroscientists are able to reveal in the brain and the interpretation of the high-level phenomena (of emotional navigation).

If we take the visual system as an example, there are cascades of sophisticated networking at the cortical level of the visual system. Striking evidence indicates that hierarchical organization of these pathways forms the basis for visual memory and spatial navigation. Integration of these attributes allows us to acquire visual perception of the complicated external world.

The same is true for the neuroscientific study of human emotions. The brain is highly compartmentalized—some parts are responsible for goal-oriented behavior, other parts—for emotions, and still others, for spiritual orientation. It is still a challenge to move contemporary neuroscience research to a higher level of understanding in terms of interactions between all neural systems that result in the emergence of such complex phenomena as emotional or spiritual navigation.

It Is Nice to Be a Human: Overthrowing Post-Humanism

The second question was again by Menon, challenging Huh's concerns about the ultimate science-fiction nightmare or a scary neuro-existentialist scenario in which human beings will suffer loss of identity resulting from mind uploading, brain-machine interfaces, and other biotechnological advances. She reminded the participants about the movie "Bicentennial Man." The movie is contextualized in the twenty-forties and it depicts a robot that is bought by a family as a companion for their daughter. At the time of the second generation the robot (the bicentennial man) falls in love with the granddaughter of the family. The robot then realizes that unless he becomes a human he will not be accepted, and he slowly goes through the process of replacing his robotic material with

human tissue. Meanwhile, he is conscious of the fact that in the process the immortal material of which he is made is being replaced with mortal tissue and, ultimately, he is going to die. In the end, humans still do not want to accept a former robot as fully belonging to the human world, and only after he dies is he finally proclaimed to be human.

Menon's illustration of an android that endeavors to become human was contrasted to what contemporary science and technology, according to Huh, are aspiring for at present. There is something nice in being a human, concluded Menon, about being faced with unpredictable challenges. In her opinion, one consequence of "boring" robotic predictability is that we would be left with nothing to solve, and hence our creativity would fade, and it is possible that eventually we would not even know how to improve our own mechanisms.

Huh countered that the above-mentioned movie is a humanist movie, and transhumanists are actually *not* concerned with going back to the human condition. Huh also emphasized that as a physician, he uses technology to help people and, even though he himself does not believe in that kind of post-humanist scenario, he believes that we have to admit that we live in a mixed culture, in a culture of popular neuroscience. In fact, his talk reflected his personal inner conflict as he is both a scientist and a Christian. This was the reason why he brought up these concerns at the present conference. Huh also argued that science has its own metaphysical assumptions that need a new metaphysical landscape which scientists may accept or not. Somehow, the Asian traditions such as Buddhism, Taoism, or Confucianism can supplement some deficits in the Western thinking, and he is looking forward to more people engaging in these debates and answering his questions in the future.

How Plausible is the Transhumanist Scenario?

The next question was by Michael Spezio, who addressed the issue of the negotiation of the world of the individual mind-brain emphasized by both speakers in their talks. Spezio suggested that some further work should be done in the area linking the transhumanist scenario to our revolutionary history in which we live. As human beings, we are extremely socially embedded, and this becomes obvious not only within religious communities, but in all aspects of life where we come face-to-face with another person, another human being, that actually defines who we are. And without that coming face-to-face, without the border, without the limit that the other person places upon us, we actually do not become who we are. In this respect, the transhumanist argumentation is illogical. In the transhumanist world we will be bored and we will have a lot of power, so we are not going to sit around and just be bored, we are going to conquer other worlds, and in the transhumanist vision, these other worlds would be other minds. So to divorce the mind from its distinct social embeddedness does not seem to be the

way to get rid of all of these problems; rather, it seems to be a way to cause even more suffering. This is because, obviously, it is not just us in this vat, there will be others around, and it will not be a virtual world, but a computational world, in which there will be others with their own wills.

In this regard, Spezio asked the speakers to comment on how they see the social embeddedness, and the real social aspect of the mind-brain-body, so that it is not just the individual property. How does it inform us? Particularly since the first speaker described his work in linking spatial navigation to, perhaps, social navigation, and spiritual navigation, including the relation with God, what is their consideration of the transhumanist vision?

Chan responded that his paradigm for neuroscientific research includes not only the body level but also the external environment, as well as social-cultural aspects. In the experiments conducted in his lab, Chan clarified, he and his team were able to address at least some aspects of interaction of the body with the external environment. They also study how certain aspects of the external environment affect the body, behavior, psychology of an individual, and their mind. In summary, neuroscientists are keenly aware of the importance of social aspects. Moreover, religious (Christian) individuals like Chan himself also recognize that interaction within the religious community is very important as well. Finally, the family component in the formation of an individual is particularly important. However, the technical problem that neuroscience faces here is that it is difficult to conduct such experiments on animals as experimental subjects.

Huh, in his response to the same question, brought up the example of identical twins, who at the time of birth have the exact same genetic material, and are completely genetically identical; they have the same home, the same parents, but in the end they develop into very different characters. This is further proof that human behavior is highly dependent on the environment. Spatial navigation, too, depends on the environment. Therefore, the transhumanist notion of a brain in a vat is self-defying. In fact, there are many sub-fields of brain science that are open to studying this issue. Social neuroscience, for example, is concerned with the study of mirror neurons and their role in constructing social embeddedness. This problematic may become a new filter for neuroscience, concluded Huh.

"Not By Brain Alone": Spiritual Orientation Unlimited

The next comment was by Bernard Senecal, who enquired of Chan if he actually conducted any studies in *time navigation*, as opposed to spatial navigation. Senecal emphasized that even though Chan earlier argued that the brain is highly compartmentalized, it is one and it acts as one. And our brains are oriented not only in space but also in time. Perhaps our spiritual GPS is oriented towards something that is Absolute Oneness—we are striving to achieve peace

with everything and with all. Spiritual traditions have found ways to orient human beings towards that peace; in the Christian tradition it is called *spiritual discernment*. People find what orients them towards peace, and they put it in practice. Senecal expressed the belief that this happens not merely in the brain or in the gut, but it is the whole body that is involved in this process. It is the whole universe that prompts the body to participate in a search for peace and harmony.

The closing question of the present discussion was raised by Paul Reasoner, who challenged an earlier comparison between uploaded minds and twins made by Huh, saying that the notion of uploading multiple copies of oneself is rather different from twins. [Ed. note: Reasoner's spouse is an identical twin.] He speculated that it would be rather different to have lived a full, rich life where one has built up many experiences, habits, and individual ways of responding to the world, and then to make a copy. This way, there is a much stronger similarity—there is much more sameness—than in the case of twins. Reasoner also suggested that it would be curious to find out what it would be like to actually meet another one of oneself, after it has already developed into a rich self like that.

In response, Huh agreed that it would be curious indeed.

Session II: Funahashi Shintarō, "Metacognition: A New Method to Study the Nature of the Mind"

The afternoon session of the conference opened with a presentation by Funahashi Shintarō, a Professor at the Kyoto University Kokoro Research Center and Kyoto University Graduate School of Human and Environmental Studies.

First, Funahashi introduced the Kyoto University Kokoro Research Center and explained that he and his colleagues at the Center define *kokoro* in very broad terms, so that it transcends our usual understanding of just the human heart-mind-spirit to also include the *kokoro* of nature, the *kokoro* of the universe, and beyond.

Funahashi specified that his research focused on human emotion (as related to *kokoro)*: how we express emotion, how we understand the emotion of others, or how emotion relates to cognitive processes. More specifically, his research revolves around the question of human self-awareness: how we understand our internal processes or how and why we actually become aware of the fact that some internal processes take place inside of us. How do we monitor our internal processes? How do we control them?

In scientific terms, this sophisticated human ability is defined as "metacognition." We know whether we remember something or do not remember it. We can consciously monitor whether we have particular information in our mind or not. Thus, the human mind has an executive function that monitors and

controls perception, cognition, and memory. Metacognitive ability allows us to perform adaptive and flexible responses. Metacognitive ability is strongly linked to declarative consciousness and language. Therefore, it has been thought that metacognition is a unique human capacity.

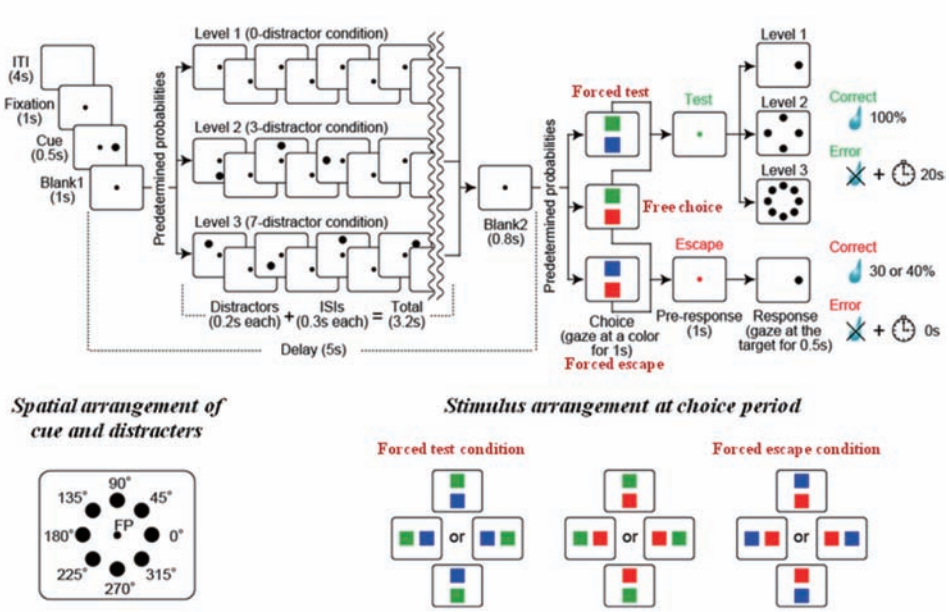
However, recent behavioral studies show evidence that non-human animals—including monkeys and birds—share functional parallels with human metacognition. Therefore, behavioral or neurobiological studies of metacognition using non-human animals should be a useful new method to understand the nature of the human mind.

Metacognitive ability has been studied in experimental paradigms that have two important features. First, the subject takes a difficult memory test which guarantees that they often feel uncertain whether they remember necessary information or not to make a correct response. Second, the subject is given a chance to handle these uncertain trials by using an escape option, in that they can escape from taking the memory test. The escape option leads to a favorable result (a small amount of reward) compared with a failure on the memory test (no reward), while a correct performance on the test yields the most favorable result (a large reward).

The basic idea that underlies these procedures is that the subject can be predicted to select the escape option in trials in which a correct answer is uncertain, if it can monitor its own memory state. Consistent with this prediction, monkeys tend to select the escape option more frequently in more difficult trials. Moreover, by selectively escaping from memory tests in uncertain trials, monkeys can improve their accuracy in a memory task, compared to when they are required to perform the same task without the escape option (forced-test condition). These observations indicate that monkeys can monitor whether or not they remember target information that is necessary to correctly perform subsequent recognition tests. Therefore, it can be concluded that monkeys have metacognitive ability analogous to human metacognition.

Although metacognitive ability was successfully demonstrated in monkeys, neural mechanisms for supporting metacognitive ability are not yet known. Therefore, Funahashi and his team developed a behavioral task requiring metacognition by adapting a spatial working memory paradigm:

After a 4-s intertrial interval, a fixation point (FP) was presented at the center of the monitor. If the subject maintained fixation at the FP for 1 s, a visual cue appeared for 0.5 s. The subject was required to maintain fixation at the FP during the cue period and a subsequent delay period. In the delay period, the difficulty of each trial was manipulated by presenting various numbers (0, 3, or 7) of distracters one by one around the FP. At the end of the delay period, the FP was extinguished and two colored squares (green, red, or blue) appeared around the FP. The subject was required to select one color by a saccade and perform a



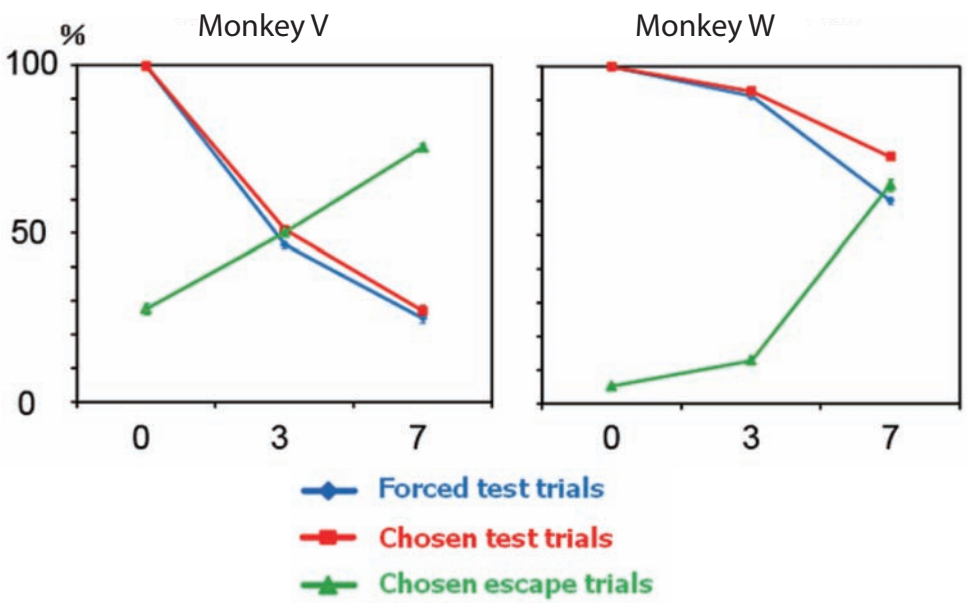
subsequent 1-s fixation. This task includes three types of choice conditions: free-choice (FrC), forced-test (FoT), and forced-escape (FoE) conditions. In the FrC condition (green and red squares), either a matching-to-place test or a visually-guided saccade task was introduced depending on the subject's color choice. For the matching-to-place test (test trial), the visual cue and distracters reappeared, while for the visually-guided saccade task (escape trial), only the visual cue reappeared. In either case, the subject was required to make a saccade to the visual cue. Correct or incorrect performance in the test trial resulted in a drop (0.5 ml) of juice or no reward accompanied by a 20-s time-out penalty, respectively. In the escape trial, only 30% to 40% of correct saccades were rewarded with a drop (0.5 ml) of juice, but there was no risk of a time-out penalty. In the FoT (green and blue squares) and the FoE (red and blue squares) conditions, the consequences of selecting green or red were the same as those described for the FrC condition. If the subject selected blue, the trial was aborted with no reward followed by a 30-s time-out penalty. First, we analyzed two monkeys' behavior (Monkey W and V) to confirm that monkeys use metacognitive capacity to perform this task. The proportion of correct performances in the FoT condition decreased for both monkeys as the difficulty of tasks increased. And, both monkeys selected the escape option in the FrC condition more often as the difficulty of tasks increased. These results indicate that the monkeys more frequently experienced mnemonic uncertainty in more difficult trials and are consistent with the first criterion of metacognition.

Funahashi's team then examined whether or not the monkeys improved their performance in the FrC condition relative to the FoT condition. Monkey W

exhibited a higher proportion of correct performance in the FrC condition than in the FoT condition. Monkey W was consistently more accurate in the FrC condition only under the 7-distracter condition. Thus, Monkey W's data were basically consistent with the second criterion of metacognition. However, Monkey V did not show any statistically significant difference. Thus, their behavioral study showed at least one monkey (Monkey W) exhibited evidence showing that this monkey used metacognitive ability during the task performance.

Funahashi explained that human neuropsychological and neuroimaging studies have demonstrated that the prefrontal cortex (PFC) plays a crucial role in metacognitive processes. However, little is known about the neuronal mechanisms of metacognition, partly because this ability in nonhuman animals is difficult to assess objectively. To address this issue, scientists recorded single-neuron activity from the dorsolateral PFC while Monkey W performed a modified oculomotor working memory task, which we developed for examining metacognition. They observed task-related neuronal activities during the cue, delay, choice, and response periods. Some of these neurons were activated differentially between trials in which the monkey chose to take the memory test and trials in which the monkey chose to escape.

In conclusion, Funahashi clarified that neurophysiologists have only recently entered this field in the search for the neuronal correlates of confidence (or uncertainty) associated with perceptual decisions and further examination of the matter is required. However, neurobiological research on metacognitive abilities using animal models is crucially necessary and important if we are



to understand human cognition. Detailed analysis of prefrontal activity may contribute to the understanding of the functional roles of the prefrontal cortex in metacognition. And this, in turn, will eventually help us understand how we monitor our own cognitive processes and how we understand other peoples' cognitive processes, their minds and hearts.

Tanaka Keiji, "Mind and Consciousness as Tools to Control Goal-Directed Behaviors"

The second presentation of the afternoon session was by Tanaka Keiji of the RIKEN Brain Science Institute, Japan. To begin his talk, Tanaka gave a working definition of *mind* as crucial for understanding his presentation. According to Tanaka, the mind is defined as the mental process consciously experienced by the subject; it does not exist continuously, in contrast to the body that always exists. The mind reveals itself in goal-directed behaviors—not in reflexes, stereotyped instinct behaviors, or habitual behavior.

Tanaka's central argument was that most brain activities occur unconsciously while they evoke actions, which we tend to interpret as our conscious decisions. For example, an anti-correlated random-dot stereogram, in which white dots in the left-eye image correspond to black dots in the right-eye image, evokes vergence eye-movements, despite the fact that human subjects never see figures, or continuous surface, there. In other cases, the mind does not know which causes actually evoked the actions.

In another experiment, when the subject is asked to select the preferred face from two faces shown side by side, they first direct their eye gaze evenly to the two faces but gradually spend more time on one face and finally select that one as preferred. Interestingly, in an experimentally-controlled condition in which one of the two faces is presented for a longer period, the subject tends to select that one as preferred. In reality, the selection was actually determined by the length of presentation, while the subject thinks that he/she selected it because he/she preferred it.

Tanaka gave more examples of experiments, which proved that our actions are also not necessarily controlled by our intentions, either. Instead, they (intentions) can be just an interpretation of an action that was evoked by other mechanisms.

A set of experiments involves rare cases of patients whose corpus callosum has been surgically cut for clinical purposes, and as a result, the left cerebral hemisphere of these patients does not have access to information processed in the right cerebral hemisphere. Because of the structure of the visual system, the stimulus images presented to the right visual field enter only the left hemisphere, whereas those presented to the left visual field enter only the right hemisphere. When the word "laugh" was presented to the left visual field of the patient, he/

she laughed. However, to the following question of the reason for laughing, the patient replied that he laughed because it was funny that the scientists conducted the boring test. In this case, the visual word enters the right hemisphere and evokes the action of laughing, but this procedure is not monitored by the language system localized in the left hemisphere. The language system instead creates a plausible reason that explains the action (laugh) in the circumstances. It often requires complex behavior to achieve a goal in a complicated environment, and it takes time to conduct complex behavior. Thus, in goal-directed behaviors, it is often necessary to maintain relevant information for a while for later use. In some cases, a set of sensory information is maintained to guide later actions. In other cases, an action plan is maintained for a later execution at an appropriate time. Such a memory function to maintain the relevant information for later execution is called working memory. The working memory requires top-down attention to select a particular piece of information relevant to the goal from always-occurring sensory input and ideas, and to protect it by inhibiting always-occurring disturbing inputs. It appears that the content of the working memory always comes to the conscious mind. Although many of the sensory inputs evoke actions and get into the short-term and long-term memories, only those that enter the working memory system are monitored by the mind. Similarly, out of many actions that are evoked by sensory inputs, preceding actions, or emotions, only those that enter the working memory system come to the mind.

Tanaka explained that the anterior part of the frontal lobe, which is called the prefrontal cortex, is essential for the working memory. Animals with bilateral lesions in the prefrontal cortex cannot perform tasks that require working memory, although they have no problems in the long-term memory.

This is also evident from the fact that patients with damage in the prefrontal cortex lose consistency in behavior. They often imitate behavior, that is, they imitate the actions of others in front of them, and perform forced tool use, that is, a stereotyped action utilizing a tool placed in front of them. This behavior occurs as if it were a reflex even when they try not to do so. Without the intact prefrontal cortex, it is difficult to inhibit the reflexive actions triggered by current sensory inputs and to direct the behavior towards the goal in the mind.

Neurons are activated in the brain to transfer signals from one brain site to others and to process information in neuronal circuitries. These processes, as a whole, control behavior. Among such neural activities, those that enter the working memory under the control of top-down attention substitute the conscious mind. Moreover, the language system interprets the behavior, probably to maintain the intention in the longer-term more easily.

In conclusion, the conscious mind emerges from brain activities for the sake of consistent goal-directed behavior control.

Discussion

Can Consciouness be Defined as "Executive Attention"?

The first comment following Tanaka's presentation was by Michael Spezio, who observed that the model of *consciousness identified with executive attention* is an inherently Western model that springs from William James's notion of perchings and flights of consciousness. This Western understanding, however, is widely criticized by practitioners of Zen Buddhist and other forms of Eastern meditation. Spezio enquired whether neuroscientists like Tanaka take into account these criticisms by Eastern meditation practitioners, and why an inherently Western model of consciousness seems to inform Asian neuroscientists more than any of the Eastern understandings.

Spezio also mentioned the critique of the "consciousness as executive attention" model by Michael Posner. [In Spring of 1997, at University of Oregon, Daniel Dennett and Michael Posner had a debate whether or not consciousness was located in the anterior cingulate cortex because that was the locus of executive attention, and Michael Posner was strongly opposed to that idea.] Does Tanaka's approach differ from this model?

Tanaka admitted that he was totally unfamiliar with the perspectives on consciousness proposed by Eastern meditation practitioners and that Eastern meditation-based critique has no influence on his research as a neuroscientist. As for executive attention, Tanaka emphasized that, in his view, the prefrontal cortex—especially the medial part of the prefrontal cortex—undoubtedly plays a leading role in voluntary action. Therefore, we can say that the prefrontal cortex is crucially important for what we understand as consciousness, but whether voluntary action is the only thing that constitutes conscious mind we do not know.

Metacognition: Brain or Body?

The concept of metacognition and metamemory, presented by Funahashi, sparked great interest among the participants and raised the most questions. James Heisig asked what role motor memory (or muscular memory) can possibly play in metacognition. Among the examples of motor memory Heisig mentioned was the body's ability to recover dance steps—we may have completely forgotten the steps but once we start practicing, our body "remembers" how to dance again. In other cases, we do not remember how to write a complicated Chinese character but once we start writing it in the air or on the palm, the motor memory comes out and we are able to write it again. Or sometimes we do not know an answer to a question but we know that we know it (as in metacognition)—it is, as we say, "on the tip of my tongue." So, is metamemory a function of the entire body? Or is it a function of the prefrontal lobe alone, as Funahashi had suggested? Furthermore, to test the above propositions, would

it be possible to design experiments with monkeys? Heisig speculated that perhaps neuroscientists could teach monkeys some complicated movements and then lead them to make the first few steps and then give them the hints and let go. If the monkeys can do the rest of it, they probably really possess metacognitive ability. Would it be possible to measure this neuroscientifically? Where (if not in the brain) can neuroscientists measure this motor memory or muscular memory? Do they even have tools for that?

In response, Funahashi admitted the delimitations of the present research in metacognition. Research in metacognition or metamemory is based on the retrospective report by language, therefore, it is very difficult to understand even in the case of humans, whether they use metonymic capacities or not. The only current method neuroscientists currently have is *the feeling of knowing judgment.* That is a very powerful method. Neuroscientists utilize a recognition memory test and collect correlation data during the judgment process in order to measure the strength of the feeling of knowing judgment. If the subject has strong metonymic capacity, strong positive correlation is observed. However, that is the only method that neuroscientists currently utilize as the third-person approach to metacognition; therefore, it is still highly questionable if there is actually a proper scientific methodology in this field.

As for the above question whether neuroscience claims that metamemory is the function of the prefrontal lobe alone, Funahashi explained that contemporary neuroscience does not make such definite claims. Currently, only a few experiments use imaging studies that suggest the prefrontal cortex activation. Therefore, neuroscientists suggest that *maybe* the prefrontal cortex participates in metonymic function—there is still very little evidence that shows that the prefrontal cortex apparently participates in metonymic capacity. But if it does, it is not clear *how* exactly the prefrontal cortex participates in it, or *what* aspects of the prefrontal cortex participate in metonymic capacities. Therefore, neuroscience today focuses on trying to find out *how* exactly prefrontal cortex participates in metonymic function based on neurophysical experiments on monkeys. There are certainly many other possibilities yet to be explored, and perhaps, "gut feeling" or some other structure, too, strongly participates in metacognitive function.

Tanaka also commented on this question, clarifying that we should be careful not to confuse metacognition or metamemory with habit—particularly, in the above example, with the dance steps that the body remembers after years of inactivity—this may be the matter of mere habit, which is the function of other brain mechanisms. Tanaka emphasized that for metamemory to occur, it has to be *monitored* by the subject, while many bodily movements we make are unconscious, such as driving, dancing, or playing tennis.

Metacognition and Self-Reflection: Is There a Difference?

Sangeetha Menon raised a rather provocative question: what is the difference between metacognitive capacity or metamemory and the unique human ability for self-reflection? How would neuroscientists define terms here? Menon pointed out that so far the discussion has revolved around metamemory in the context of task-oriented behaviors. However, most human behaviors are not simple one-to-one task-oriented behaviors. Human experience involves various degrees of reflection: we make choices, comparisons, we set priorities and ascribe different values to different things not purely based on our metacognitive capacity—human behavior is a complex phenomenon. So, what is the difference between metacognition and self-reflection, if any?

Metacognition and Self-Awareness in Religion: "The Eye Cannot See the Eye"

A similar question came from the audience: what is the difference between metacognition and a religious concept of self-awareness? A religious concept of self-awareness refers to being aware (or not being aware) of a present state of consciousness, rather than that of the past (as in metamemory). Apparently, it is easier to be aware of what one did last year (week, day, minute) than to be self-aware in religious terms of the current moment, or, as we say, "the eye cannot see the eye." Perhaps, the hard case for the study of metacognition would be the study of self-awareness of the cognition of the present?

However, is self-awareness the same as self-reflection, or are they two different phenomena? Menon objected that primates, perhaps, may be *self-aware*, but only human beings exhibit complex levels of self-reflection. In other words, self-awareness is clearly related to task-oriented behavior that is a functional possibility, of which other non-human primates are capable. Self-reflection, however, is *not* a task-oriented behavior. It includes making choices, prioritizing, being emotionally sensitive, being creative—all this is possible because of the human capacity to self-reflect.

The above discussion was summarized by Frank Budenholzer, who suggested that the gap between the human and animal sense of self-awareness is defined by self-objectification: in human self-reflection one becomes the *object* of self-understanding or self-analysis.

"Why Did I Go Upstairs"? Metacognitive Capacity and Aging

How about the age of the experimental subjects? Do neuroscientists take this into account? With age, there is an increased tendency to forget things, or as Yamamoto Sukeyasu put it, "You go upstairs and you forget why you went upstairs."

Menon and Funahashi commented here that *the ability to remember that you are not being able to remember* what you had planned to do is another kind of metamemory. In fact, metacognition is a broad concept that includes t*he ability to know that you do not know something.* Or, again, is this a kind of self-reflection—not metacognition? The question remains.

Metacognition in Animals: A Myth?

The final question on metacognition was raised by Newsome, who challenged the concept behind Funahashi's experiments in metacognition on non-human primates. Can we really be sure that the monkeys have metamemory, or meta-cognition? Do they truly possess this subjective "feeling of knowing" or a feeling of confidence about their memory? Can we really decipher the intrinsically first-person nature of consciousness based on experiments with monkeys? Or, is it possible that through months and months of training the monkeys have learned the certain sequences of stimuli followed by certain choices that have a higher probability of a reward?

Funahashi argued that the monkeys' score in metacognition-related tests is higher than the score in the forced test conditions, and this pattern is continuous. This is especially clear from the experiments that have an "escape-option." If the monkey knows that there is an escape option, it will always prefer an "escape-option" to a "test-option." This proves that monkeys have a meta-cognitive ability. However, ironically, the same very fact may imply that the monkey is actually reward-oriented, as long as he interprets an "escape option" in terms of a reward. Is metacognition in animals just a myth?

In Top-Down Control, What Is at the Top?

The final question was addressed by Newsome to Tanaka, and it challenged the numerously-mentioned notion of "top-down control" in Tanaka's presentation. What is at the top in top-down control? What do we really mean by this? Neurophysiologists and psychologists use this term all the time, but what do they really mean by it? What is at the top? Is there a level of organization at the top that exerts downward control? Or is all causality by definition upward because it is all about physics?

In response, Tanaka clarified that he uses the term "top-down" to mean "from pre-frontal cortex to motor or to sensory cortex." He specified that for him pre-frontal cortex is "at the top," not because it has some super-neurons but because it has a connection with all parts of the brain. In this case it is more "global" or superior.

Newsome, however, argued that the real challenge for neuroscientists today is to create a new terminology or a new language that will allow talking about the reality of top-down causality in terms other than merely "from one cortex to

the other." "I want to talk about these collections of states of neural systems as having causal efficacy governing the activity of neurons at lower levels," insisted Newsome. Is it possible to define goal-oriented behavior as a constellation of neural states that exert downward control from higher levels to the organism? If yes, is this even science? Or is it philosophy? Where and how can we talk about it? The question remains.

Session III: William Waldron, "No Brain is an Island: The Intersubjective Construction of Experience in Buddhism and Cognitive Science"

The final presentation of the day was by William Waldron, who presented a critique of current cognitive theories of religion through the prism of classical Indian Buddhist philosophy of mind. Current cognitive theories of religion claim that the reason people attribute agency to "supernatural beings" is due to specific cognitive modules in our brains, such as a Hyper-Active Agency Detection Device (HADD). The main focus of Waldon's critique was the assumption that the relevant cognitive processes can be adequately explained in terms of modules in each individual's brain, a classic "brain-in-the-vat" theory. He argued instead that complex human cognitive processes, such as the attribution of agency, can be better explained by patterns of causal interaction that have come about in both human evolution and personal development (phylogeny and ontogeny). As Waldron demonstrated, both cognitive scientists and Indian Buddhists have gradually shifted from individual-based models of cognition to more intersubjective models.

Specifically, most cognitive scientists have heretofore assumed the primacy of the individual and looked for causal mechanisms exclusively in an individual's brain. Some cognitive scientists now suggest, however, that our distinctively human "worlds of experience" arise through intersubjective cognitive processes that are based on culture and shared symbolic reference (language, and so on). Our tendency to attribute agency, for example, arises out of such a nexus of causal influences. But since these processes occur mostly unconsciously, their influences are difficult to discern, and it is the task of science to disclose and analyze these hidden causal patterns.

To clarify his point, Waldron cited theories of linguistic constructivism, which holds that distinctively human cognitive processes are in large part a function of language, or, more precisely, of the linguistification of the human brain. Since language, as we know, refers to reality *virtually*—not directly—we can conclude that our sense of *self as agency* is virtual, too. To take this idea further, supernatural agency is but another kind of virtual reality.

Waldron demonstrated that Indian Buddhist theories show remarkably similar developments. Early Buddhists also initially analyzed cognition solely at an individual level. Later Buddhist thinkers (4–5th c. CE) expanded the range of relevant causal influences and argued that our "shared worlds of experience" (*bhājana-loka*) depend upon intersubjective cognitive processes, most of which are, however, largely unconscious. They argued that language—the medium of human communication par excellence—is crucial for the construction of common human experience. The attribution of agency is generated by largely hidden, yet deeply intersubjective, causal influences.

As evident from the above, in order to better explain the genesis of human meaning and experience some cognitive scientists and Indian Buddhists are looking beyond "brain-in-a-vat" models of individual cognition and are articulating intersubjective models that include the influences of language, society, and culture. In conclusion, our individual sense of self is not bounded by our individual mind or body, and we cannot even begin to understand ordinary human consciousness without appreciating the importance of sociality of it. "No brain is an island."

Michael Spezio, Response

No Discipline is an Island: The Urgency of Interdisciplinarity

Michael Spezio introduced himself as a Presbyterian minister, working on the borders of science and religion, where representatives of various academic disciplines come together to make meaning of the questions such as those discussed at the present conference.

In his response to Waldron's presentation, Spezio reconfirmed that just as much as "no brain is an island," no academic discipline is an isolated entity, and cognitive science is a good example of it. Cognitive science is a highly interdisciplinary field. As George Miller (Department of Psychology, Princeton University) demonstrated in his paper, "The Cognitive Revolution: A Historical Perspective" (*Trends in Cognitive Sciences,* 2003, 141–44*),* cognitive psychology, anthropology, computer science, information science, and neuroscience are all part of cognitive science.

In unison with Waldron's challenge to rethink the importance of sociality in the way we construct our cultural worlds, Spezio strongly emphasized the urgency of interdisciplinarity today. We need interdisciplinarity not only between the fields related to cognitive science, he insisted. Neuroscientists today cannot do without theologians, philosophers, scholars of religion, and so on. Interdisciplinarity, however, does not imply that we should depart from the disciplinary rigor of our respective academic fields. On the contrary, it means

taking the methodological rigor of all these disciplines most seriously, for this is the only way to have the most highly mutually challenging, yet truly constructive dialogue. As evident at the present conference, not only all participants from the field of neuroscience demonstrated this kind of disciplinary rigor in their presentations, but Waldron's presentation also brought the same kind of rigor from religious studies.

Tripartite Explanatory Level: Giving Voice to "That Thing in the Scan"

Spezio identified as one of the running themes throughout the day *the notion of explanatory level*, with human experience being its ultimate focus. He also argued that if *human experience in all its fullness* is the ultimate goal of neuroscientific research, then we cannot be constrained to a third person perspective as the only safe domain of conversation about causation. Traditionally, scientific interpretations of causation (and the present conference is no exception here) takes a third person point of view: we are trying to get an idea from outside what they—that person in the scanner or that monkey in the chair—are doing. However, it is evident that as soon as we take a non-third person perspective, or when we actually take *a second person* perspective, this is when we actually acknowledge that the experimental subject has deeper dimensions to them, whether it is a monkey or a human. It is when we are confronted by the possible reality of the Other that we go deeper than our surface models approach and begin to do serious interdisciplinary work. William Newsome's presentation on experiments with monkeys doing direction discrimination tasks was a vivid example of this second-person approach: when the monkey makes the wrong choice in a task (artificially forced by means of electrical stimulation of his brain cells), the question here is not whether it is possible to manipulate the monkey into doing something by stimulating his brain—the question here is what the monkey is really realizing at the moment. And it is a much deeper question. Is the monkey saying "Was that not the correct answer? Why did I not get my reward?" Or is the monkey saying "Darn it! Why am I doing it?" That is when the second person arises and, of course, then we bring back the third person, and the monkey becomes simply a subject again. There are ways to try and tease out those two possibilities, not by any means technical, but very clear ways.

Spezio defined this kind of all-embracing approach as *tripartite nature of causality,* which includes perspectives from the first person, the third person, and, yes, from "that thing in the scan"—the second person, who is actually always there at a potential go-to position. It is when we actually realize how socially constituted every one of us is; even a monkey in a chair has had a kind of social environment, and even a monkey raised away from its mother has had the social environment of its human captors. Perhaps this tripartite methodology can help us rise to a more profound, all-inclusive explanatory level.

Human Experience: Prediction vs. Explanation

Spezio also addressed the importance of distinguishing prediction from explanation in scientific approaches to human experience. Simply put, what can be proved right in the laboratory may not work at all in real life. For example, there are well-known models from human reading, such as that developed by Gary Feng (Department of Psychology and Neuroscience, Duke University), and the "E-Z reader" model developed by Keith Rayner (Department of Psychology, University of Massachusetts). These models predict 90–92% of the time not only where people look when they are reading a text, but how long they spend actually focusing on the words. These predictions are astonishing. The problem, however, is that they are built with a false assumption that how long you spend on something is completely independent from where you go to next, as demonstrated in the works by Scott McDonald (Center for Cognitive Science, University of Edinburgh). These models can be right 92% of the time in terms of where people read, but they are wrong as soon as the conditions are modified.

This is an example of why neuroscientists and cognitive scientists have to be very careful in terms of explanation: a certain series of behavioral events may well be predictable within a controlled laboratory setting; however, it gets far more complicated under different conditions or in real life situations.

On the one hand, we need disciplinary rigor in conducting experiments. On the other hand, we have to be aware that once the content changes, this particular prediction may actually be fooling us as to how correct our model is. This is why the notion of *external validity* is so important to us.

If laboratory science done on isolated individuals can absolutely call into question whether humans are free, whether human consciousness is efficacious for behavior, whether humans are actually moral or altruistic—if scientists can call those things into question without any external validity provided by other disciplines, then they might actually be making a grave mistake. It is impossible to approach human experience in all its fullness without generalizing on many different contexts or without generalizing too many complex situations in which agents are socially embedded. And this is why we need—once again—an interdisciplinary interaction.

The Centrality of Sociality to Neuroscience

Responding to Waldron's challenge to reconsider the significance of sociality for the understanding of individual consciousness, Spezio clarified that social psychology and social neuroscience are currently employing new methodologies that place individual consciousness into the context of social interactions.

One of these is the *social brain hypothesis* that says that the brain as we have it is actually strongly influenced by the social contexts that shape it. Another is the

notion of *joint action*, which is a brand new field (in fact, it is one of the newest fields in the area of social psychology and social neuroscience) that is trying to get beyond the individual in a scanner by putting her into interaction with other experimental subjects. This allows measuring with a high degree of precision what contributes to the efficacious coordination of that action by both subjects involved in a joint action, be it a competition, such as tennis match, or a poker game, in which deception plays a major role. Indeed, there are many questions that cannot be answered by studying individuals in isolation.

Finally, there is another interdisciplinary approach that neuroscientists (Spezio included) employ for the understanding of consciousness, human freedom, et cetera. It involves experimental work with exemplars in the moral domain. These exemplars have to be empirically verified, of course, however, neuroscientists approach these exemplars not only from the third person perspective, but also from the first and the second person perspective, allowing subjects to speak about their own experiences. In doing so, neuroscientists learn from their experimental subjects how to design appropriate experiments, identify appropriate measures, and so forth.

Topdown–Topdown Causation

Spezio's final comments addressed Tanaka's arguments on the importance of understanding the role of the subliminal in goal-oriented behavior. Spezio argued that while neurosciences—system neuroscience, cognitive neuroscience, social neuroscience, and social psychology—all recognize the strong influence of the subliminal, this implicit influence is actually termed within these fields as "top down." Despite the fact that this influence is subliminal, it is still "top down" because it comes from organized structures, conceptual schemas, and it is not determined by simple stimuli coming from the outside, but it is dependent on developmental history and perhaps on evolutionary history of psyche as well.

Spezio identified this top-down subliminal influence as having a great potential for further studies. As he put it, "We do not only worry about top down–bottom up, we have to worry about top down–top down." It is top down—top down in terms of being consciously aware of the kinds of implicit processes while trying to interact with these highly structured schemas within the mind, instantiated or at least contributed to by patterns within the brain. Spezio pointed out that these themes have been previously discussed by Warren S. Brown and Nancy Murphy (Fuller Theological Seminary) and many others.

In conclusion, Spezio once again emphasized the importance of recognizing the tripartite nature of the human for the explanation of human experience, and the need for strong and really intentional interdisciplinary work in the future.

Open Discussion

As at the previous international conferences on "Brain Science and Religion" in East Asia organized within the GPSS project, the main foil to the Buddhist perspective in the discussions was Sangeetha Menon, representing a 2500-years-old classical opposition between Buddhist versus Hinduist (Vedantic) understandings of human continuum.

Four-Dimensional Self: A Vedantic Perspective

Addressing Waldron's presentation, Menon admitted the importance of approaching a complex concept of "self" or "human experience" in a dynamic context. She explained that the classical Hindu tradition recognizes the significance of inter-subjectivity and the importance of the *second person approach* for reducing the tension between the first-person and the third-person in both scientific and philosophical terms.

However, she questioned whether it is legitimate to entirely reduce the concept of "self" to a mere social construct (linguistic, social, and developmental). Even if we admit that self is experiential—it is formed in interaction with society—can we equate "self" with *only that?* Perhaps, it will be too hasty because self can also indicate something that is more ontological, a space that is not yet defined, but by which we are all inspired, and we are continuously inspired to think and see something that is undefined and that we want to reach.

Menon suggested that it may be useful to understand "self" as a multidimensional phenomenon, as depicted in the ancient Sanskrit text called *Mandukya Upanishad*. *Mandukya Upanishad* presents the concept of *turiya* or the fourth state of consciousness.[3] Menon speculated that the difference between the first- and the third-person, or the role of the second-person in intersubjectivity can be enhanced if we also have the concept for a fourth-person or state (what in Sanskrit we call *turiya*). Thus, a Vedantic perspective offers a deeper, more inclusive dimension of "self," it helps us understand "self" in all its richness.

Finally, Menon challenged Waldron's notion of the cognitive and experiential limitations of self. According to Menon, it is when we realize that we are situated in a limited context that we try to go beyond the limited, to enhance our mechanisms, to enhance our experiences, to enhance the richness of self. Realization of our limitations is what should push us, should launch us to do something, to break through, to change, to develop—our identity is not completely exhausted

3. The *Mandukya Upanishad* (1-2 cc. AD) describes four states of human consciousness, namely a. *susupti,* the dreamless sleep state with *prajna* as the experiencing self; b. *svapna*, the dream sleep state with *taijasa* as the experiencing self; c. *jagrta*, the wakeful state with *vaisvanara* as the experiencing self; and d. *turiya*, the fourth conscious state of super wakefulness, which is an extraordinarily awakened, self-integrated state. The primal being or essential awareness in such a state is called "atman the essential self." [A.G.]

by the limitation of our cognitive and experiential mechanisms. This gets added significance in Eastern meditation practices, which enable us to see the limitation of the limited and prompts us to go beyond it.

The Buddhist Critique of Esssentialism: "The Truth Will Set You Free"

In Waldron's counter-objection, the conception of trans-experiential self may be a mere question of semantics. Hinduism suggests the notion of the trans-individual dimension of "self," which in a sense is "above experience," but if there is indeed such a self, then what work does it do? *What philosophical work does it do?* How does it evolve if it is not the part of causal processes?

Buddhism considers the notion of self as an agent (or as something transcending temporal experience) problematic because it invites the problems of property dualism. A classic Buddhist objection to that kind of concept of self—and Buddhism has a tendency to side with science here—is that we have to deal with the world that is causally closed and our task in terms of the dialogue between science and religion is to understand the mechanisms of this world, whatever they may be, to understand their causal history and their causal properties. Once we understand the causal mechanisms of the world we live in, we can get free from these limitations.

The social phenomenon of nationalism can serve as a good example here. On the one hand, when we recognize that nationalism, or a sense of national identity, is a social construct, and as such shaped with political ramifications, it actually frees us from its power. On the other hand, nationalism (or national identity) is still a very powerful concept—people tear up when they hear a hymn of their country, it is a deep emotional experience. And yet, it is not ontological; it is something socially constructed. In this sense, cognitive science and neuroscience play an important role in society because what they do is they free us from our constraints: the causal patterns by which the world runs become revealed through scientific analysis and we are not slaves of our unconscious mechanisms any longer. For example, yes, hormones are powerful but still we are free to choose. Cognitive science and neuroscience open for us an entirely new realm of autonomy. And in this sense the realization of our limitations is very important indeed.

Is the Buddhist Critique of Reification of Essences Still Relevant?

The participants also questioned the significance of Waldron's critique of the unchanging self in the contemporary social context. They objected that, certainly, from a neuroscientific point of view—even from a common sense point of view—we are not unchanging selves. Synapses in our brain change, we change with experience, we grow and develop, we change over time, and in this sense,

how is this traditional Buddhist critique of the unchanging self still relevant to us today?

Waldron clarified that his critique of the unchanging self is dual. First, it is a historical critique of the concept of *atman* originally developed in the Upanishads, the early set of Hindu texts from roughly the sixth to fourth century BC. In the Upanishads *atman* is described in terms of an unchanging essence that comprises the core of a human being, and it is presented as the highest ontological fact. It is translated into English appropriately as "self." Buddhist thought, or the historical Buddha Siddhārtha Gautama, rejected such a notion of an unchanging essence as being the ontological foundation of reality and offered instead a conception of self as a unit of experience, which is ongoing, evolving, and perpetually changing from moment to moment. Thus, Buddhist thought has been constantly in opposition to Hindu philosophy for the last 2500 years until the present day.

We may mistakenly think that the concept of the unchanging self is an anachronism; however, even today we still have this unconscious commitment to the concept of the unchanging self. We get into the fallacies of thinking that there is some fundamental core in us that does not change. In fact, there are many positions within Western philosophy as well that will say that a self or consciousness has to be in some way immaterial.

How is the Buddhist critique of self relevant to neuroscience? Waldron argued that any vibrant philosophical religious tradition has something to offer to the ongoing discussion and can enhance research in cognitive science and neuroscience. Therefore, Buddhist arguments are not going to be formulated

in a way that they were 2,500 years ago, but there are certainly perspectives and modes of analysis that are continuously relevant.

Waldron's critique of the trans-experiential concept of self was further developed by the participants who addressed both the Yogacara's historical critique of the unchanging self, and the later works in Zen Buddhism and other Buddhist traditions, which attempt a deconstruction of the emotionally-bound sense of self. Our ego gets hurt when we are offended or confronted on the personal level, and by deconstructing the notion of self Buddhist philosophy helps us emotionally eliminate the pain we may experience. In this sense, Buddhist thought is also meant to set us free (help us get detached from this world's passions)—on the emotional [heart] level as much as on the mental [mind] level.

Is There a Post-Deconstructionist Self?

Bernard Senecal objected to Waldron's critique by stating that even if we deconstruct the classical ontological notion of self as an unchanging essence—simply put, even if we say that the self is not what we think it is—it does not necessarily mean that there is no self. Maybe there is a different kind of *self*?

In response, Waldron paraphrased the Buddha's words, "If there is an unchanging something that is the self, show it to me." What is a self? When we use words like "mind" or "self," do we refer to an actual thing that is separate from the complex context in which it is used or in which it may arise? Where can we find this context-free self? This is why we can say that deconstruction of linguistic referentiality lies at the core of the Buddhist critique of the concept of self.

Different passages in early Buddhist texts present philosophical deconstruction of the possibility of ontological essences being involved in the causal world: "Where is it? Let's focus on what there is." If there is an ontological essence, where is it in relationship to experience, to feeling, to thoughts, to memory, to our bodies, which are transient and always changing? Where is it to be found?

Instead, Waldron argued, the basic question Buddhism asks is, "Why does suffering arise?" Buddhism is concerned with the causal mechanisms of the world, and if something like an unchanging ontological essence is not a part of the causal world, Buddhism cannot answer this question. The problem of why something arises may not be resolved by recourse to ontological categories.

I Feel, Therefore I Am

Menon strongly objected to the idea that self is but a product of semantics—a linguistic and processual-developmental unit. In her argument, self is *experiential,* first and foremost: "I feel hurt, I feel loved, I feel jealous, I feel passionate…." We can explain the concept of self linguistically but we can know it only experientially. We can exhaust the concept of self within the linguistic paradigm, where

we see self more as an assembled, relational phenomenon which is prone to deconstruction. But there seems to be a possibility to explain that "something" exists in an experiential realm, and not merely in a linguistic realm.

I experience something today and *I* experience something tomorrow; things change but there is still something essential, which seems to be abiding in *me*, which *I* know, and how do *I* know that? Because *I* am able to form certain values, priorities, certain likes and dislikes, evaluations, and choices. So, it seems that we cannot make evaluations or set priorities unless there is something that remains as an essential core. Perhaps self is not a completely individuated entity, but there is still a possibility for a self that is a core of our being, something which is undefined, but which allows us to look beyond.

Menon emphasized that there is another powerful argument in support of the idea that there is an essential core in the human psyche. She brought up the examples of neuropsychiatric disorders, where the self is continuously shifting, and the boundaries of the self blur. In spite of all that, however, the patients are capable of manifesting positive qualities: willpower, love, empathy, and so forth. So, evidently, there is something deep inside their being that helps them to come out of their mental confusion and make some sense of their existence. In fact, it can even help the patients make sense of their own neural and mental impairment. So, again, we should not easily abandon the concept of an essential self, because it opens up a space from which we develop and get positive qualities: compassion, love, empathy, and others, which make us human.

Western Critiques of Anti-Essentialism and More

Frank Budenholzer commented that, while many mistakenly believe that the conception of the unchanging essences is an inherently Western (specifically, Aristotelian) idea, this is a false assumption. Aristotle never talked about essences in a sense of unchanging entities. He recognized the changing nature of essences, yet he developed the concept of the basic whatness of things, the whatness that fundamentally stays there and does not change.

According to Budenholzer, a more profound Western critique of anti-essentialism comes in the form of a question: "If there is no unchanging self behind our beliefs, if we are mere streams of consciousness, then how do we know that what we believe is true? If there is no *I*, how do *I* know my science as a scientist? There seems to be something that does not change and allows us to go forward. Essentialism also lies in the basis of much of our ethical thought. We still feel that the whatness of essences has to be maintained and referenced somehow. If a person has lost most of his or her abilities to experience, we still believe that this person should be provided care. Ethical problems concerning stem cell research, abortion, euthanasia, and so forth, revolve around the question of what it means to be human—what it means to be *essentially human*—not in a

sense of unchanging entities but in a sense of the whatness of a human being, human dignity. In Buddhism, we can play with the notions of self/non-self but this challenge of the whatness of human existence does not go away. There is a far deeper problem involved.

The participants further challenged Waldron's anti-essentialist critique by pointing to the fact that—paradoxically—he consistently utilizes the term "object of awareness" in his explanation of the Abhidharma Buddhist notion of human continuum as "thoughts without a thinker." According to Waldron's explanation, awareness is constituted of an object of awareness, faculty, and attention. However, the question that arises here is, "If there is an object of awareness, is there not *a subject* of awareness?"

Evidently, the empirical self can be deconstructed. We can strip the self off all socially constructed particularities: "I am x years old, I have a college degree, I have a family, and I am a professional"—we can dispense with all of that, it is non-essential. While it is fashionable to talk about "thoughts without a thinker," the idea that there is an *object* of experience without a subject of experience seems self-contradictory and is hard to understand.

Waldron pointed out that the Buddhist response to this is that we cannot intelligibly separate subject from object except for analytic purposes. They are both part of an ultimately inseparable process.

What Is the Whatness of the River?

Waldron classified all of the above criticisms as purely semantic misconceptions. What is the whatness of the river?, he asked in response. What is the essence of the river that does not change despite the accidental quality of the water, the accidental quality of its currents, of its banks? Are we are looking for a river that does not have any water, that does not have currents, does not have a river bank? What is a river beyond its constituents? What exactly would that be? What are we looking at?

We can say, perhaps, that the river is outside, while the self is inside, and therefore this metaphor is inappropriate, but does it change anything? We are using the word "self" here as the same kind of category. We are acting as if there is going to be a referent there, something essential remaining after we have taken away all the transient temporal processes of experience. If we take away all of the changing processes of the river: the water, the currents and the river banks, is there still going to be something there, some kind of abstract riverness? Where is that river? There is no river; it does not have a universal referent. It might have a referent in terms of various categories of thinking about the world, but if we try to look at the datum there, there is nothing to look for in terms of the riverness: it is just the water, the current, the gravity, and so forth.

The same principle lies at the basis of the Buddhist anti-essentialist critique of self: we have feelings, we have emotions, there is continuity, our experience changes radically from one moment to the next—as a result, we have an illusion that we are an ontological unit. But if we abstract or extract everything experiential from it, and what do we have in the end?

Buddhism speaks of the self in terms of the constituent categories of existence. And since we use language to refer to the world, we operate in terms of linguistic categories. Eventually, what we end up doing is unconsciously essentializing or reifying the most general categories that we use. As a result, we assume that behind these categories, there is something essential, uninstantiated. We make ontology out of linguistic references.

As Karl Popper puts it, from the scientific perspective categories are normal and necessary tools that we use to understand phenomena in all their complexity but it doesn't mean that we have to impute an ontological existence into everything for which we have a name. Otherwise, the world becomes endlessly populated because language is constantly changing. In the end, reification of essences becomes a *new kind of polytheism.*

Self as a Unicorn

The final comment was made by Alena Govorounova, who pointed out that Buddhist anti-essentialist critique may evoke some serious moral and ethical implications by defying the concept of personal responsibility. A criminal's brain or a criminal's self is embedded in social networks, and following the Buddhist logic of interdependent arising, when a criminal commits a crime, every single unit of the same social network is responsible for the crime to an equal degree. How about individual responsibility?

A typical Buddhist response to this challenge, as Waldron put it, would be that to deconstruct "self" is like deconstructing a unicorn: we assume the reality of what we are asking about and the Buddhist response to such a question would be, "Let us define terms." As for the practical dimension here, the Buddhists would rather ask, "how can there be any type of criminal responsibility or any other responsibility if the self actually were unchanging, since, if a person were *truly unchanging*, there would be, by definition, no cause and effect relationship pertaining to such a self?" And the Buddhists are primarily concerned with the notion of causality—that there are effects of one's actions (karma). Hence, the Buddhists argue that there is personal responsibility precisely because there is no unchanging self. Self is a unicorn.

Session IV: Sangeetha Menon, "Brain-Challenged Self and Self-Challenged Brain"

The central argument of Menon's presentation was that the significant problem in consciousness studies today is perhaps not the notorious Chalmersian "hard problem" of mind-body dualism, but the question of how to trace the ways in which the brain challenges the self, and the self challenges the brain. That is to say, how do the brain and the self conceive their roles and create the conspiracy of experience whereby the physicality of the brain is lost in the subjectivity of the self? Based on examples from current research in brain studies, neuropsychology, and neuropsychiatry, Menon proposed a theory that just as there are neural correlates of consciousness, there are also *self-correlates of consciousness* (both positive and negative) that seem to alter the functions of the neural correlates and "challenge the brain."

To start with, Menon gave working definitions of the concepts of the brain and the self. The brain is one of the most important parts of the human body, which is today studied to understand the working of sensation, emotion, and consciousness. The single unit of information and experience that connects sensation, emotion, and consciousness is agreed to be the "self."

According to Menon, there are two major streams of discussion. First, self is considered as a cognitive concept that helps tie the missing ends between the physical and psychological functions, and second, self is argued to be the seat of conscious experience. However different the arguments for these two positions are, it is agreed that human behaviors, attitudes, and emotions are intricately tied to the neural structures on one side, and the indivisible experiential self on the other. Brain and self are the common threads that are used by neuropsychiatry, neuropharmacology, and philosophy to make some claim on one of the most intractable problems of humankind, namely, "consciousness."

Is there a common issue in brain and self studies that appears over and again? Yes, namely the attempt to explain the unity, continuity, and adherence of our experience, whether it is sensory or mental. To address these experiences is to address the place of the self in the brain. A major challenge to this effort is the fact that, though we tend to commonly address a static unit by calling it "self," the self is constantly emerging as a result of its interaction with nature outside (social and biological) and nature inside (mind). In the process of its emergence the boundaries of self seem to change, creating havoc for some (in the case of misidentification syndromes) and peace for others (in the case of spiritual experiences).

Philosophically we continue to ask the question about mind-body unity, how the mind and body—with their different natures—can connect and give rise to meaning and quality to life. The binding problem and the Chalmersian "hard problem" showcase the age-old mind-body problem in the context of

consciousness. Both demand mechanisms and reasons for mutual influence. The interconnections between brain and self have been especially eschewed in the developments in understanding the brain and its functions. The classical idea about the brain with designated cortical areas and assigned functions, though, is no longer in vogue; the view that supersedes it is that the brain is an organ with a high capacity to survive, even with less cortical areas. There are medical cases where patients seem to live "almost normally" in spite of a frontal lobotomy or cortical lesions due to psychiatric conditions (Todd E. Feinberg, 2001, *Altered Egos: How the Brain Creates the Self*). It is suggested that perhaps the limbic system, the seat of emotion, is the most important part of the brain without which normal functioning is impossible (Antonio Damasio, 1994, *Descartes' Error: Emotion, Reason, and the Human Brain*).

The brain also has the capacity to switch a neural function to a different area if the designated cortical area is impaired; this capacity is termed "neuroplasticity." The plasticity of the brain is accompanied by yet another enigma—that the self is somehow able to make sense of the neural changes and create corresponding changes in sensations and personal identity. Menon suggested that just as there are neural correlates of consciousness, there are also *self-correlates of consciousness* (such as compassion, love, quietude). Self-correlates (both positive and negative) seem to alter the functions of the neural correlates in curious ways.

Where and how in the brain is the "self" housed? How does the self make adaptive changes in the person corresponding to changes in the brain? How does the self influence and alter neurochemical functions of the brain? Can the brain address its structural and functional challenges without recourse to the self? Can there be a self without the interface of the brain and the limbic system? Are the brain and self constantly challenging each other?

These and similar questions may not give an immediate answer considering the complex ways in which both our brain and self is wired to our concepts and causal thinking. We do not even have many different ways of understanding the subject and the object other than by causal relations. The medical cases studied by neuropsychiatrists often show that the way the patient behaves before and after a cure is not even amenable to conclude that there are straightforward causal relations between brain and self. The subject-object distinction itself is violated when the brain behaves in ways not true to its essential physical neural structure. Can the brain be called distinctly objective and physical when it defies the laws of medicine?

Menon concluded that it is important to continue the classic mind-body debates, but it is equally significant to understand the emergence and placement of self in the context of an evolving brain which has the capacity to be plastic. Greater insights into the nature of self—neural and ontological—will arrive if we focus our research on the challenges that the brain and the self give to each other.

Discussion

Science and Religion: Pushing Each Other's Limits

As the discussion began, James Heisig challenged Menon by saying that her presentation echoed age-old debates on the relation between mind and body, where the word "body" has been replaced by the word "brain." How the body (brain) becomes conscious, how consciousness is embodied—this is typical mind-body rhetoric. Obviously, neuroscience has not been able to explain the human self in all its fullness. But is there anything that research in brain science or neuroscience can possibly discover, any conclusions that the scientific community can possibly formulate that will allow us to reform—not simply enhance or reconfirm—but reform ideas that have been circulating in the perennial philosophies for centuries? Maybe if we take a closer scientific look at self, it will turn out to be a mere fiction? What assumptions about the classical doctrines of self are we—philosophers and theologians—prepared to let go of, or to change, or at least to seriously reconsider in the light of what brain science has found out? Has there been no essential reform?

Heisig also suggested that, instead of focusing on the weak points of brain science and on the unanswered questions, maybe it is more constructive for religionists or philosophers to engage in the scientific process rather than resist scientists by fixing on the limits of science? Otherwise, it almost seems like religious believers heave an obvious sigh of relief over scientific limitations. As was evident from Kyoon Huh's earlier presentation, Heisig argued, neuroscientists themselves are keenly aware of the limits of brain science and the (transhumanist) dangers entailed by scientific attempts at the deconstruction of the concept of self. Neuroscientists recognize their own limitations but they see them not as a way to stop thinking or avoid challenge but as an agenda to push their own limits further. What would be religion's response to that?

Focusing on the weak points of neuroscience, Menon agreed, is of no use, for the limitations of neuroscience is a given. It is delimited by definition. Neuropsychiatric literature gives us many examples of how the fragility of self is a mystery for neuropsychiatrists as much as it is for philosophers. By recognizing the limits of neuroscience we are *not* painting a negative picture; in fact, it is good sometimes to heave a sigh of relief at the understanding of one's own limitations. Menon emphasized that rather than engaging in faultfinding with neuroscience, she sees her calling in bringing into the dialogue the importance of human *experience* of the self. Mind-body debates and the like, abstract speculations—philosophical or neuroscientific—will never end, but understanding *experience* is a real challenge for us as scientists to understand. This is why, Menon insisted, she ascribes such high value to self-correlates such as love, compassion, empathy, and other experiential constituents of self. Self

as we experience it is no less important for our dialogue than any third-person scientific explanations. Menon concluded that it is very important to be a skeptic after all; knowing one's limits is the first way to understand how to cross them.

Self as Experience

Is self a mere fiction, then? Menon reconfirmed that, indeed, the concept of self is often critiqued—her own work included—in terms of a social construct: self is created either linguistically, or synaptically, or neurally, or as a narrative. These discussions are ongoing and they are endless, too. However, what we always forget to bring into the discussion—and what remains her main concern here—is the concept of self that is familiar to all of us: the concept of *our* self, something that is totally unique to us; something that is known only to ourselves; self that *we* experience. And it is crucially important to engage with such a self. Even more importantly, there may be no progress in the understanding of self unless we have a passionate relation with what we are trying to realize.

Spezio and Menon further discussed the range of delusional misidentification syndromes in which patients deny the existence of their paralyzed body parts, or falsely attribute them to a deceased family member, or say that a dislocated body part "is not obeying today," and so forth. Menon suggested that these phenomena may be explained by the brain's capacity to create metaphorical meanings in order to make sense of the oddities that happen. Metaphorical representations help patients reconstruct their understandable experience of self.

Is There a Self Beyond Language?

Commenting further on the discussion on metaphorical representations in the brain, Govorounova inquired if self may be associated purely with the brain's linguistic ability to make sense of experience. Does that mean that self is housed—or is being constructed—in the left brain hemisphere? Most importantly, may there be some kind of essential, non-linguistically-constructed self? Is there a self *beyond human language* and, if yes, what is that?

Is self located in the left brain hemisphere? According to Menon, we do not know: perhaps, it is located in the body, perhaps, all over the body or even outside the body. Menon eventually defined self as "something that seems to be influencing the brain." Otherwise, self can refer to an ability to make coherent connections between many discrete, unrelated experiences.

Huh Kyoon commented on this discussion by saying that self may be also defined as an ability to have control over bodily movements. Thus, split brain patients, for example, do not have an ability to make sense of their experience linguistically because their left brain hemisphere is damaged. Split brain patients often have a problem controlling their movements and their hands might "fight each other" as these people are trying to perform a task. However, curiously

enough, split brain patients are *aware* of their problems with movement control and they make conscious efforts to make harmonious movements. Evidently, there is a sense of self-awareness that springs from the right hemisphere of the brain and it is probably not correct to say that only the left brain hemisphere may produce self-reflective experiences. Self is beyond language.

Bernard Senecal, "Neurological Underpinnings of Zen Meditation and Christian Spiritual Exercises"

Senecal introduced himself as a Jesuit priest who for the last two decades has been specializing in the study and practice of Korean Sŏn (the abridged form of Sŏnna, the Korean transliteration of the Sanskrit word *dhyāna*: meditation, which has become *chan(na)* in Chinese, *zen(na)* in Japanese, and *thien(na)* in Vietnamese). Senecal defined the main thesis of his presentation as the problem of the inner conflict that Christian practitioners of Zen meditation experience with their doctrinal tenets, and challenged the audience to consider whether neuroscience or cognitive science may be helpful in bridging these two religious traditions by explaining the neural basis of mystical spiritual experiences.

Senecal has encountered many Europeans, Koreans, and North American Christians who, as they practice Zen, end up experiencing unfamiliar states of consciousness that they cannot easily reconcile with the basic tenets of their faith. As a result, while some decide to abandon their Zen practice, others choose to stick to it in order not to lose the benefits that they get out of it, even as they keep experiencing difficulty in articulating what is happening within their minds.

What is this inner doctrinal-experiential conflict that Christian practitioners of Zen meditation experience and how can it be described? Senecal defined it as a "long-lasting and distressing impression that language has become disconnected from reality and that action has become meaningless." This impression is distressing not because it takes place during meditation but because it persists even after meditation is over. This impression, of course, can only be achieved as a result of deep and longitudinal practice of Zen meditation (Vipassana retreats can have the same effects).

This state of consciousness does not impair a person's ability to speak, hear, read, or behave and express oneself normally. There is nothing abnormal or unnatural in the behavior of the person experiencing it. The problem, however, lies in the fact that the person experiencing this state of consciousness cannot get rid of an uncomfortable mental impression that nothing that he does, senses, hears, reads, writes, and even what he thinks means much anymore. It is a feeling of being disconnected. Senecal argued that this fact implies that there is something in our brain or in our body that originally (prior to meditation) makes us feel that we are related to the phenomenal world and to our selves.

Meditation, thus, has modified the brain in a way that a person does not have a sense of connection to his self and the world.

Christian traditions, obviously, have developed their own meditative practices, thus, for Christians practicing Buddhist meditation it should not be problematic to "take off in the plane" of their consciousness into a mystical realm and to safely "land" or return to habitual reality at the end of meditation. However, the problem for Christian practitioners lies in the fact that they know how to "take off," but they do not know how to "land"; it is an unfamiliar realm and they do not recognize the landscape any more. They go beyond the realm of language where they can encounter Absolute Nothingness and they get lost. They have lost their compass, they do not know who they are any more. However, since Sŏn meditation is embedded in Buddhist metaphysical underpinnings, Buddhist practitioners do not seem to have a problem "landing." Senecal concluded that Buddhist and Christian meditative techniques are only superficially similar—at core there are significant differences. The fact that some Buddhists make scathing criticisms of so-called Christian Zen, in which they see an amalgam of utterly contradictory doctrines, or that the hierarchy of the Catholic Church has kept warning their flock against the risks of Buddhist meditation, argue in favor of the seriousness of the case.

Although Buddhist-Christian encounters can be dealt with at a great variety of levels—mystical, philosophical, theological—Senecal insisted that the time has come to examine these phenomena more specifically, and the question today is whether neuroscience can help us to understand what is the exact nature of the conflict that may take place in the brains of Christians practicing Zen, and whether there are ways to overcome it or not.

Discussion

Trans-Experiential "Pure Consciousness" Experienced

The discussion began with a question by Govorounova inquiring if a practitioner of Buddhist meditation can experience a sense of self which is not merely "beyond language," but also beyond experience as we normally understand it.

Senecal shared that, indeed, during his meditative practices he occasionally experienced moments of "pure consciousness" that were not only "beyond language" but also beyond normal human experience of one's self as a coherent existence. These flashes of "pure consciousness," however, may last only during very short periods of time. To illustrate, Senecal gave a metaphor of climbing Everest, where a climber can stay at the top of the mountain for a very limited period of time; otherwise it will be life-threatening. Then, a meditator will have to "land" or return to reality and be able to experience again the habitual

continuity of experience, or the experiential self that Sangeetha Menon had previously described. However, Senecal insisted, there is no experience that can pretend to be disconnected from interpretation. Therefore, even the experience of "pure consciousness" is still inevitably tied to language and expressed through language.

Menon commented that, in her opinion, as soon as we start talking about self in the context of language, it brings us experientially to the context of "me and the Other." And the true experience of self (in essentialist terms) is the dissolution of "me" and "the Other" where language does not exist. At that very point, where there is no Other, one can experience "true self" in the true sense of being. The function of language, thus, is to report experience, but language itself is not equal to experience, and we should not confuse the two.

Inputs from Neuroscience

Funahashi Shintarō commented that, according to his research, memory plays a crucial role in identifying the continuity of the self in a given individual. Working memory (short-term memory) plays an important role in the formation of the sense of self-awareness, as does the interaction between working memory and long-term memory.

Huh Kyoon suggested a neuroscientific input on the development of better "landing" techniques for practitioners of Buddhist meditation. He described fMRI scans of the default mode of the brain. Default mode of the brain means that experimental subjects supposedly do nothing during the scan process: they do not think, do not meditate, they simply rest. The purpose of such experiments is to analyze what parts of the brain are activated when the brain supposedly does nothing.

Huh also explained that according to research by Andrew Newburg (Associate Professor of Radiology and Psychiatry in the School of Medicine at the University of Pennsylvania), during meditation the activity in the temporal lobe decreases but it increases in the frontal lobe instead. Huh's suggestion was that in order to "land" or return back to reality after meditation, a meditating subject, neuroscientifically speaking, should come back to the default mode of the brain, to his regular resting condition.

Senecal strongly objected to this suggestion. His basic argument was that once the subject has experienced this kind of deep meditation, his "resting condition cannot be the same anymore." In other words, there is no regular default mode to come back to; it no longer exists. And unless the subject finds new meaning of his existence and finds a new way of being at peace with oneself and the surroundings, this is not going to work.

Senecal also argued that, in his experience, both the brain and the body are continuously looking for harmony of existence—this is the basic driving force

behind all our actions. True spiritual experiences aim at the sensation of peace and harmony with the world, and it can be experienced on different levels.

The participants further discussed the future potential of research in metacognition of spiritual experiences. James Heisig suggested that, perhaps, research in metamemory or metacognition can inform us of the nature of the higher states of consciousness, whether they are stimulated by drugs, or by electrodes, or by some kind of spiritual reflection. Is there any way to measure the metamemory of those experiences neuroscientifically? Is it possible to measure the memory of someone recalling the experience of "taking off into a mystical realm" and "coming back to reality"? Is that too introspective for neuroscience to study or is it possible at some point? These possibilities are yet to be explored.

Satō Tetsuya challenged the participants to take Senecal's research agenda of the neuroscientific study of meditation onto the deeper level. In Satō's argumentation, the purpose of Zen meditation is to disconnect from human noises and external signals in order to connect with the intrinsic spiritual reality of nirvana. The purpose of Christian meditation is to experience the power of God on a deeper level. But what is the purpose or the necessity of the neuroscientific study of these spiritual practices? What kind of new information do we expect to get?

Senecal explained that his main agenda is to understand from a neuroscientific point of view how and why human beings are programmed to enter into another realm of reality. Why does one have this urge to explore these higher states of consciousness onto the end? Simply put, a frog living at the bottom of the well needs to get out of the well to see that there are other wells. And in order to do that it has to get to the source of its own well first.

Satō insisted that in the context of the present conference, the most important questions are whether God's power really exists and whether nirvana is a "real place," so to speak. Is it possible to discover the power of God through detailed examination of the brain activities? Perhaps, neuroscientists can do a comparative research on Christian and Buddhist practitioners and discover the differences between their spiritual experiences of God versus nirvana? Maybe neuroscience may end up discovering that there is a God or there is a nirvana—or the absence of those. Otherwise, why is this research necessary at all? How does it help us?

In response, Senecal reminded the participants that, as Christians, we are in front of the Great Mystery and even neuroscience cannot put God's power "under the microscope." Neither can it exhaustively explain nirvana. But, referring to Chan's notion of the "spiritual GPS," Senecal suggested that what neuroscience *can* do for us is help activate our inherent spiritual capabilities. This does not mean that neuroscience of the future will produce "spiritual GPS machines" and send them to us in the mail. But it can teach us how to better use our own

"spiritual GPS" which we all have but often do not use because we do not have a "manual."

Senecal also mentioned that there is another, rather humbling, dimension to this problem. As Zen philosophy teaches us, "Do not try to understand—just be." Enlightened beings are one with the Tao, in the sense of transcendence and immanence. Buddha, Christ, and great sages of the past were not concerned with the neural basis of spiritual experience. But they all—in contrast to us—knew well how to use their "spiritual GPS."

Neuroscience of Spirituality: Setting Reasonable Expectations

The final comment of the present discussion was made by William Newsome, who suggested that we should be careful about what is reasonable to expect of neuroscience and brain science in regard to research in spirituality. In response to Senecal's challenge to neuroscientists to examine spiritual phenomena and help religious believers enhance their spiritual experiences or resolve inner conflicts, Newsome warned that these expectations might be unrealistic, and clarified that they probably stem from a common misunderstanding about neuroscience as a field. Newsome insisted once again that the central dogma of neural science is that all parts of mental activity emerge from, and are deeply linked to, the activity of the brain. This means that neuroscience studies behavioral observations and seeks to explain neural changes that various experiences cause in the brain. If the behavioral changes are real neuroscience can ultimately find brain areas that correspond to them. If neuroscience can trace some real reproducible neural changes in the brain that result from spiritual experiences, it can certainly study them. However, while neuroscientists can potentially understand

the mechanisms of spiritual experiences, it does not mean that this understanding will lead to *an enhanced experience* on the part of spiritual practitioners.

A good example of this is a neuroscientific theory of color vision. Knowing the theory behind the color vision does not really enhance our experience of color, it does not change our experience of color. However, it explains some illusions of color and it helps us make color machinery, such as cameras, or printers. But just as neuroscience does not change our primal experiences of color, it might not be able to enhance in any way our spiritual experiences, even if we come to know about their neural basis. On the other hand, Newsome admitted that neuroscientific study of spiritual experiences can enhance our understanding of the laws or mechanisms by which they happen and satisfy us in our spiritual search on the explanatory level: all of us who have spiritual experiences want to know what is going on inside the brain—or are we just fooling ourselves? Unfortunately, at the present level of neuroscience some spiritual phenomena are still too complicated to decipher, simply because we do not have the tools for that yet.

In the end, Senecal expressed hope that the day will come when neuroscience will be sufficiently equipped to conduct such research. Meditation affects the brain to such a degree, argued Senecal—that one day somebody will have to elaborate a new interpretation of this phenomenon and explain to us how it affects our neural connections.

Session V: Iriki Atsushi, "And Yet It Thinks…."

Why There is No "Mind Science": The Non-Reproducibility of Mind

Iriki began by referring to the so-called "central dogma of neuroscience" previously discussed by William Newsome in the opening presentation, that all mental phenomena are based on brain activity, and thus, the conceptions of mind and self are the mere products of the neural activity of the brain.

Popular public opinion, however, remains that while mind and brain activity are obviously tightly linked together, causality works the other way around: "*kokoro* moves the brain," that is, brain activity is caused by the functioning of mind. This was evident from public comments on a government report, *How to Promote Brain Science in the Future,* submitted recently by the Brain Science Council of the Japanese Ministry of Education, Culture, Sports, Science and Technology. During a public session in reaction to this report, a portion of the respondents too large to ignore seem to claim that they do not believe in upward (body to mind) causality and they do not accept the notion that all mental activity is a mere product of the brain.

Ever more interestingly, Iriki pointed out, most scientists in basic science (molecular neurobiologists in particular) claim that "mind" is not a subject of neuroscience and should not be studied based on brain function. Why is this the case? Why do many scientists believe that mind is a non-scientific category and cannot be analyzed on scientific grounds?

The main reason, according to Iriki, lies in the fact that the concept of "mind" does not fit into the conventional paradigm of reproducibility, universality, and falsifiability that defines the natural sciences. Indeed, there is no such a thing as a "reproducible mind." Every mind or self is absolutely unique and cannot be reproduced. In addition, something that constantly changes its inner rules or its status does not fit into the traditional paradigm of science either, making it too hard to formulate the fundamental rules governing "self."

Non-Measurability of Mind

Another reason for the non-existence of an official "science of mind" is that the functions of the symbolic system in the brain are too difficult to measure technically. While non-symbolic sensory and motor systems in the brain are tightly linked to actual reality and their functions can be easily traced, the functions of the symbolic system do not correspond to anything tangible or objectively present in the real world.

For example, we can scientifically analyze the function of our visual system: when we see the color "red," five hundred nanometers of electro-magnetic waves that compose the "green light" activate neural retinal cells in our eyes and then the signal is carried onto the sensory cortex of our brain. We can correlate the activity of the sensory system to the physical entity (electro-magnetic waves) and we can actually measure this activity. The same is true for any observable behavior; we can actually measure movements, forces, and physical parameters of behavior and try to relate them to the nervous system's activity. No wonder, then, that these "hard core" sensory and motor physiologies have been predominant in neuroscientific analysis until recently: they fit into the classical paradigm of *the scientific method*.

However, when neuroscientists observe that the association cortex area in the brain lights up when humans appreciate justice, or love, or beauty—what can they do then? Obviously, these symbolic notions must be activating some neural cells in the brain but with what can neuroscientists correlate them? How can they scientifically define or physically measure *beauty* or *justice*? It is impossible to correlate abstract notions with the nervous system activity and there is no way to formulate the rules by which these correlations occur. This is the reason why the human symbolic system of representation has not been the subject of neuroscience until recently. Science by definition deals exclusively with the physical world and its observable phenomena.

Biology: A Historical Quest for Recognition

The final explanation for why there is no "science of the mind"—in Iriki's analysis—may be found in the history of biology and the fact that biology had to undergo a historical struggle for recognition *as a science* approximately one hundred years ago. "The Empire of Physics" dominated natural sciences at the beginning of the twentieth century, while biology (or *a science of life*) was seriously questioned as a discipline. Life is something unique, individual, non-reproducible, and constantly changing, so the question was: how can biology fit into the classical scientific paradigm of universality, reproducibility, and falsifiability? It was not until molecular biology appeared and justified biology as having real chemical-physical grounds that biology was recognized as a proper natural science. Iriki suggested that this may be the reason why, historically, biologists in general and neurophysiologists in particular are reluctant to swim in the risky waters of scientific uncertainty. Nobody wants be flushed back to those times when they were not recognized as real scientists and, therefore, nobody wants to deal with something as ephemeral and spooky as "mind science."

If "Mind Science" Violates the Definition of Science, We Need to Change "Science"

"And yet it thinks," insisted Iriki. Paraphrasing Galileo's famous claim, Iriki asserted that the concepts of mind and self remain crucially important for us as legitimate tools by which we define our relationship with the world, both scientifically and mundanely. There is an undeniable need today for making "mind" the subject of scientific inquiry, and it cannot be any longer ignored. As William Newsome had previously demonstrated in his presentation, the issues of "mind" and "self" are no mere philosophical speculations—we have arrived at a point in history where natural scientists have started routinely facing these philosophical questions in their laboratories. "Mind" has to be a part of the scientific dialogue of today and if "mind" violates the definition of what science is, then, maybe, the time has come for us to revise the four-hundred-year-old definition of "science."

As a first step in including "mind science" into classical scientific enquiry, Iriki proposed to define "mind" and/or "self" as a *hypothetical explanatory concept*—an *illusionary* notion that does not have a referent in the phenomenal world but can be utilized as a means to explain our relations with the world in terms of causality. To explain this idea, Iriki drew parallels between the concepts of "mind" and "self" with the concepts of "god" in religion and "gravity" in physics. "God" and "gravity," according to Iriki, are not real objects—they do not "reside" anywhere but they can be used as convenient tools for the explanation of causal sources and relationships in the natural world.

However, the real question here is: how can we accomplish this major paradigm shift that will allow us to include "mind" and "self" into the framework of scientific method? How can we modify the unchangeable rules of classical Newtonian physics? Iriki suggested that one way is to revise our understanding of "science" as restricted by the principles of "universality," "reproducibility," and "unchangability." Iriki argued that one way to overcome our rigidity and narrow-mindedness in regard to the definition of science is to consider the fact that the universe itself and its laws are not as universal, reproducible, and verifiable as it seems at first glance. The universe itself keeps creating many new things and many new rules that did not exist before.

For example, the rules governing the primordial stage of the universe immediately after the Big Bang are different from those we experience now. In the beginning, the universe was governed by high energy physics or particles physics, probably due to the super high primordial temperatures. In the beginning, there were no chemicals yet, only particles floating in a super-hot primordial "soup." Once the universe cooled down, chemical reactions began, producing different materials. As a result, the rules of chemistry appeared, followed by the rules of Newtonian physics that governed the relationship between different materials and physical objects. As we can see, different sets of laws and principles were coming into existence at different stages of the development of the universe: first, the laws of chemistry emerged, then the laws of Newtonian physics. Finally, four billion years ago biological life emerged and the laws of biology appeared—in other words, the universe constantly keeps birthing new phenomena and new laws and principles, and our world is not as "unchangeable" and "reproducible" as it seems.

Iriki concluded that currently we are at the stage where we have to approach "mind" or "self" as a kind of a "mini-universe," which constantly creates new concepts and thoughts, governed by new rules, schemes, and laws. And while "mind" or "self" does not have a physical correlate in objective reality, it appears to be an extremely powerful and influential force, which constantly transforms and renovates the world in which we live. Iriki insisted that for us to feel comfortable, we *need* the concepts of "mind" or "self" in order to explain this constant creation of new metaphysical and physical dimensions that we perform by our mind-force.

How Can We Research "Mind" in the Lab? "Self" in Hydra and Protozoa

In addressing the practical potential of the scientific study of mind and self, Iriki began with a fundamental question: why is there a mind? Putting it more practically, why is there a brain? Why would living organisms present on Earth have a nervous system? Obviously, the function of a nervous system is to sustain life—but what does it mean to "live"? Iriki suggested that to live, first of all,

is to overcome entropy: life involves falling apart, and yet, it is simultaneously striving to organize itself back, preserve itself, and keep the cells of the organism together by means of the consumption of energy.

Not all living organisms have nervous systems, however. Plants do not have a nervous system, because they are grounded and do not move physically in space. But almost all animals that move physically in space (with the exception of jellyfish)—that are "animated"—possess nervous systems because they function within the predator-prey behavioral scheme: they need to find food, hunt, identify their enemies, and escape from their enemies. This is how animal organisms evolutionarily differentiated motor apparatuses from other parts of their bodies and developed sensory organs.

Iriki argued that a *configuration of self* is what defines animal life: even the simplest single cell organisms like protozoa or hydra have a *configuration of self*—not the concept of self but a physical configuration of self. Simple organisms are closed physical entities that are surrounded by membranes allowing them to have physical boundaries and perform input and output of energy and information. As the organism structure became more and more complex, central nervous systems and brains emerged in higher animals in order to process information more efficiently. So, it is important for animal organisms to differentiate themselves from others and from their environment, and it is crucial for us to finally recognize the significance of "self" in biology.

Human versus Animal Intelligence: What is Special About the Human Brain?

So, what is the human brain? What is the human self? Why is it so special? How is the human brain different from the non-human brain, and how is human self different from the basic *configuration of self* in space? How can we technically study the most highly organized forms of mental activity?

Iriki clarified that animal intelligence is composed of a sensory system (sensory cortex of the brain) and a motor system (motor cortex of the brain) that allows animals to adapt to the external world. Human intelligence, however, also includes a symbolic (linguistic, semantic) system of representation: it allows humans to create metaphysical concepts detached from the physical world. But the association cortex is not exclusive to humans: animals (specifically, non-human primates), also possess the association cortex. So, why is there a difference between human and animal intelligence? What is it that allows humans to create the world of metaphysics, which transcends objective reality?

In order to account for the volutionary emergence of human intelligence, Iriki proposed a classification of self-identification in living organisms into *intransitive* and *transitive* categories. Iriki claimed that all animals, except for non-human primates, process information and control movements *intransitively*: they do not sense or perceive themselves as more than motor or sensory

apparatuses. Non-human primates, however, are different: they possess an elongated hand, which allows them to use tools and perceive their tool-equipped extended hand as an object, and that implies a possibility for primates to perceive themselves *transitively*.

Laboratory Experiments on Self-Body Awareness in Non-Human Primates

According to Iriki, his curiosity as a neurophysiologist prompted him to wonder: "Why not conduct scientific experiments on the association cortex of non-human primates? Why not teach monkeys to use tools that would allow them to develop a symbolic system of representation and make their non-human brains more human-like?" To test this proposition, Iriki's team conducted a set of experiments training monkeys to use tools (a rake) to catch bait projected onto the TV monitor. In this experiment, a monkey is not directly looking at the bait but he is looking at the TV monitor, which shows the image of the bait and his hands holding a rake captured by a camera. The monkey's real hands are blocked from his view by a black screen, so the monkey cannot see his hands' movements directly; instead, he is looking up at the TV monitor and is trying to grasp the object with a rake by watching the movements of his hands projected onto the TV monitor. Arguably, the monkey in this experiment is projecting himself (his self-body image) onto the TV monitor. In the same experiment, if the scientists project onto the TV monitor a scary figure like a spider or snake next to the monkey's hand, he will immediately retract his hand with an expression of fear, which means he must know that this is his hand on the TV screen. So, there must be neural correlates that neuroscientists can measure to identify the mechanisms of the coding of the self-body image by integrating sensory (visionary) and motor information. There are neurons that respond to the tactile sensation in a monkey's hand and there are neurons that respond to the visual information about his hand without tactile sensation. Neuroscientists define this as a receptive field. By measuring the neurons that project a receptive field onto the monitor and the neurons that code the subjective self-body image, neuroscientists can study neural activity corresponding to the subject's self-introspection.

Self-Objectifying Ability in Non-Human Primates

Iriki also argued that once a monkey is trained to respond to tools, the neurons in his brain start responding equivalently to his hand and to a tool, meaning that the brain starts perceiving a hand and a tool as having equal status. In other words, in a monkey's mind, a tool is incorporated into a hand or, alternatively, a hand becomes extended (with a tool). Iriki suggested that the same evolutionary principle lies at the basis of human development from birth to maturity: a newborn baby interacts with the environment mainly through tactile sensations;

as it grows older it develops iconic or visual perceptions, and finally, an adult person has a fully developed symbolic (semantic, linguistic) system of representations. IRIKI argued that non-human primates do not have a symbolic system of representation yet, but they may have the potential for it.

Based on this assumption, Iriki's team conducted a set of experiments trying to artificially expand the range of non-human primates' intelligence. How is human tool-usage different from that of animals? Non-human primates presently can use only mechanical motor tools. In contrast, humans also use external sensory devices, which upgrade or extend human sensory organs, such as a telescope, microscope, microphones, x-ray machines, radio detectors and many others. Moreover, humans recently started using "external brains" or "memory storage devices"—computers, the internet, and so forth. IRIKI suggested that there is an evolutionary hierarchy in tool usage development: perhaps our ancestors first acquired the ability to use motor tools, then external sensory tools, and so forth. Once the ability is acquired, it becomes a default function of the brain operation and the species is ready for the next step. So, the first step on the path of humanizing the intelligence of monkeys would be to try to train monkeys to use external sensory devices.

Iriki's team has trained monkeys in the above-described experiment to use an endoscope-like camera (instead of a TV monitor) in order to grasp the bait. This way, the monkey was using an "externalized eye" to trace the movements of his "externalized hand" in trying to reach the object. Interestingly, while it took only two weeks to teach the monkeys to use motor tools, it took nearly three years to teach them how to use sensory tools. IRIKI argued that in this experiment the monkey perceives himself through an "externalized eye," which is a third-person perspective or self–reflection of a kind.

More specifically, in the monkey's mind his hand becomes equivalent to the tool, and he starts perceiving his hand as an object. This is the key moment where the important transition happens in the monkey's brain: self starts observing oneself as an external object. Also, here is where a hypothesized explanatory concept of self is formed in the brain—a necessary conceptual link that explains this self-objectifying, self-projecting behavior of oneself to oneself. And since we know that logical connections in the human brain are established in a pre-frontal cortex, we can say that this is where our "self" emerges. But neuroscientifically speaking, what would be the neural correlate of this? Iriki's team conducted another set of experiments using fMRI and discovered that the *temporo-parietal junction* area of the brain is responsible for this kind of projectional activity, where the subject objectifies oneself or reflects upon oneself. Again, "self" here does not refer to any substantial object and it does not reside anywhere—it is simply a *hypothetical explanatory concept*. Iriki concluded that

the origin of the concept of self in humans evolutionarily can be traced to *transitive* movement, that is, manipulating objects and tool-usage.

A Real Challenge to Reconsider What is Real

In conclusion, Iriki once again highlighted his principle ideas: first of all, the brain is a living organism that constantly creates new concepts and rules, and the creation of these new concepts and rules necessitates some kind of hypothetical explanatory concept equivalent to gravity in physics or god in religion. This hypothetical explanatory concept of self may be a useful tool for the account of the emergence of consciousness: without self or intention, living organisms passively adjust to the environment. With the emergence of self (intention, goal-oriented behavior), however, living organisms can intentionally act upon the environment and adapt to the environment. As they adapt to the environment, they start using new tools, which results in the expansion of their brains, which consequently means more advanced tool-usage, and then, further expansion of their brains and so forth.

Our brains create unique, irreproducible, constantly changing worlds, but our mind-self has not been the subject of conventional science until recently because conventional science is exclusivist and rigid in denying everything beyond its grasp. But currently we are facing the challenge of a new paradigm shift where we might have to expand the conventional boundaries of science. Many things in this world do not fit into the conventional paradigm of universality, reproducibility, and falsifiablity, but this does not mean that they are not physically real.

How can we explain why an airplane flies? It flies in physical reality, but there is no physical causality of flying that is known to us. We know that the speed of the passage of the air is different between the downside and the upside of the wings of an airplane, and that when this speed difference occurs, the lift of the plane is produced. However, nobody can really explain the physical causality behind these two events—there is simply no official scientific explanation of why an airplane flies. And yet, it flies.

In the very same way, mind-self does not have any tangible physical correlate in actual reality; it does not "reside" anywhere, nor does it refer to any substantial object. And yet, it thinks. It is real.

Discussion

The participants widely challenged Iriki's ideas both on technical scientific and on conceptual levels. What is the difference between the neuroscientific concept of *self-projection* that is the key notion to the explanation of the emergence of human consciousness and a basic human capacity to empathize with others?

Neuroscientists say that humans *project* their self-image on others and they can also *project* others' actions or emotions onto themselves, thus expanding their subject-object or subject-subject paradigms. But is it possible that all the above is no more than a mere ability to empathize with others, and maybe neuroscientists are simply fooling themselves? So, what is the difference between self-projection and empathy, and is it possible to neuroscientifically identify it?

Also, the participants questioned the above-described experiment where the monkey uses the rake to catch bait and, specifically, Iriki's claims that the monkey identifies the rake as an extension of his hand. Is it true that the monkey considers the rake to be his extended hand, or maybe he is able to distinguish his hand from the rake? In the above experiment, Iriki demonstrated that when the monkey sees a scary figure like a spider, snake, or scorpion, he withdraws his hand in fear. However, if the monkey would use a rake or another tool to push a scary figure away or to kill it—which he would never otherwise do with his hand—this would mean that the monkey clearly knows the difference between his hand and the tool. So, can we really say that in the monkey's head there is no difference between his hand and the tool?

The most vivid discussion revolved around Iriki's main message that the present definition of science must be revised and expanded. If universality, reproducibility, and falsifiability are not good enough criteria for defining science, then what is? On the one hand, we have to agree with Iriki that there are many legitimately scientific phenomena that are unique, irreproducible, and unfalsifiable. In fact, evolutionary theory would be the most representative example of this. Evolutionary theory cannot predict anything; it deals with unique and unreproducible phenomena all the time—in fact, we can almost apply the principle of "predictable unpredictability" to evolution. Interestingly, there are many natural scientists—among them some very famous physicists and chemists—that do not believe in evolution because they do not consider it scientific. (Nevertheless, some participants argued, evolutionary theory is still commonly accepted as science because within it there is some fundamental continuity between the basic physical, chemical, and biological laws. Secondly, evolutionary theory is not a personal narrative, and in this sense it can also be considered scientific.)

The discussants also argued that another powerful support in favor of revising the present scientific paradigm may come from the history of science, which demonstrates that the definitions of falsifiability, reproducibility, and so forth, have never been stable or absolute—they have changed historically, reflecting cultural and intellectual climates of given periods. This implies that the fundamental criteria for defining science have always been flexible and, in fact, political.

Psychology and medical sciences are even more "in trouble" with the present definition of science than evolutionary theory because they have to deal with personal narratives all the time. And in this sense psychology and medical sciences have the potential to bridge the gap between experiments and experience, and reconcile personal narrative with science.

So, what next? If we include personal experience into the scientific framework, what will this mean for science? More specifically, what does it mean for the science and religion dialogue? Will we be able to include personal experiential narratives about individual spiritual experiences (from New Age and esoteric traditions, and Zen meditation) as legitimate counterparts in the dialogue with science? How can we verify personal spiritual experiences like meditation or supernatural revelations?

The participants agreed that there should be reproducibility in these experiences at least in the statistical sense. For example, in medical sciences, the experience of pain is a personal narrative but it can be verifiable statistically. So, personal experience must be *sharable* for the overall scientific understanding or mechanism to be discovered. There must be some kind of statistical distribution of personal narratives for us to be able to create a precedent so that we could fit personal narrative into the current scientific paradigm.

In conclusion, the participants agreed that the definition of science is ambiguous; however, presently the conventional paradigm of universality, reproducibility and falsifiability seems to be the only legitimate one for doing science. It is crucial for us to have some conventional verification methodologies to define scientific methods, otherwise we will have to go back to Aristotle and interpret natural phenomena speculatively. And yet, there seems to be a looming paradigm shift, threatening to shatter our most basic understanding of what science is. This challenge stems specifically from the present dialogue between science and religion, and we should be prepared to adequately respond to it. How we deal with this challenge and what kind of fundamental paradigm shift we may experience in the near future is going to be the next chapter of the dialogue between brain science and religion.