

Hitoshi NAGAI (Nihon University)

Why Isn't Consciousness Real? (1)

Preface

In the commonly held world-view, each person has what is called 'a mind' or 'consciousness', and it is produced mainly by a person's brain activity. Even those who do not believe the second half of this would believe the first half. In the following series of lectures, I would like to demolish this world-view. However, I will not simply demolish it. I will elucidate the mechanism through which it must have been established. I will demolish it in the sense that I will show it to be a fabrication, but I would also like to show that that fabrication is indispensable to us, and that it constructs our 'actuality'.

Thus, the three lectures will have a dialectical structure which is not summarizable in the form of a set of theses. They will be dialectical in such a way that a thesis affirmed at one stage may be denied at a later stage, or a thesis denied at one stage may be affirmed at a later stage. Moreover, the discussion will not unfold systematically towards a truth, but will rather construct a fiction as 'actuality', and will even form a loop. The meanings of the words used will change accordingly.

Let us take the concept 'zombie' as an example. A zombie is a creature that is indistinguishable from a human externally, but is not conscious internally. As our discussion progresses, the claim 'I am not a zombie, but others are', the claim 'Because zombies are conceptually impossible in the first place, neither I nor others can be zombies', and the claim 'Zombies are possible, and even I myself could be a zombie' will each be affirmatively asserted as an indispensable, indubitable truth. Furthermore, it will be clarified that the first claim, 'I am not a zombie, but others are', for example, has more than one meaning, and that therefore the same can be said of the claim 'Zombies are possible, and even I myself could be a zombie'. This method of discussion *per se* will be presented as inevitable and indispensable for the understanding of the concept of a 'zombie'. Since a 'zombie' is, by definition, a creature lacking 'consciousness', this discussion will apply, as it is, to 'consciousness'.

I hope that the lectures will be *read* carefully with the above point in mind. Although I would admit the clumsiness of my style, and some possible slips of the pen (or tongue), I believe that the truth of the thought expressed in this book is inviolably definitive.

The lectures are based on the seminar I gave at the Faculty of Letters of Osaka University in the summer of 2006, and at the Faculty of Humanities of Niigata University in the summer of 2007.

Day 1: Why Is Consciousness a Philosophical Problem?

There is nothing general to be called ‘the mind’.

The mind, especially what is called ‘consciousness’, does not, in fact, exist. That is a self-evident fact that anyone knows. When I say this, most people are bewildered. Even professional philosophers who take an interest in this kind of problem, or indeed materialists who usually insist that the mind is only a function of material things, start to say such a thing as, ‘But shouldn’t we admit that our own mind, or our own consciousness, exists?’, which rather surprises me. Such philosophers, in most cases, presuppose the existence of a mind or consciousness, and are wrapped up in such inconsequential topics as how the mind or consciousness relates to functions of material things such as the brain or nervous system, which I find rather disappointing.

If you will forgive my impertinence, I would say that professional philosophers are not actually doing philosophy at all. They are merely following the rules of a ready-made game, a game of unknown origin which has arisen spontaneously, in which they all agree to follow the same rules. Whenever they play this game, the result is ultimately unsatisfying because they never question why the game itself exists.

It is often said that the mind or consciousness is in fact produced by the brain and nervous system. Such a problem about the nature of the relation between the mind and the brain – or, more generally, the problem about the relation between the mental and the physical – is called ‘the mind-body problem’. But for me, this *problem* is extremely puzzling. I can never understand how it is possible to formulate the topic in that way. And I want to find a way to understand how it is possible to formulate it that way.

In my view, it does not matter which position one takes after the problem has arisen. One could happily pick randomly from such positions as mind-body dualism and physical monism (i.e. materialism). Any difference between these positions is not essential, and it becomes even less so for some of the other positions. Whether we are to acknowledge the origins of the problem in the first place, and, if we do, how we are to explain these origins, is all that is philosophically meaningful.

I have never ever seen anything called a mind, yet it is believed that there is such a thing generally – that is, for me, you, him, and her. Isn’t that the problem to begin with? Isn’t that more puzzling than anything else? Why don’t people think that there is a problem here? In addition, what do people mean by ‘material things’ or ‘physical things’? Isn’t it true even of the brain that we identify it by perception? Before we discuss the relationship between the brain and the consciousness it produces, should we not first discuss the relationship between the perceived brain and the physical brain? With these problems left out, any discussion of

the relationship between the mind and the (physical) brain is a castle in the air. I cannot even understand the meaning of the question.

As I have said, I have never seen anything you might call a 'general mind'. You might then ask: Does that mean that one only sees – or feels – one's own mind rather than the 'general mind'? The answer is 'no'. I have never seen or felt anything like 'the mind one oneself has, which is not general' and is felt by each person. I can only feel my own mind. I am sure that it exists. Yet I can never know the existence of anything as general as 'oneself'.

Since there can only be one instance of it, my mind is not general. Furthermore, because that sole instance *is and can only be* mine, is it not enough to say 'mine' rather than 'my mind'? Is it really necessary to say 'mind' if the idea is that there is no such thing as the mind? Of course, this could be put conversely: It is also enough to say 'mind', such that 'my' is not necessary. Either way, it remains the case that there is only one instance of 'my' and 'mind'. Therefore, it might as well only be called 'this'.

In spite of my views, why does everyone else believe that there is something we can call a 'general mind'? That should be the very first problem to be elucidated. If it is missed out, philosophy of mind is not philosophy and would mean that most of philosophy of mind cannot count as philosophy. The solution to this problem is in the following three-part lecture. Or I should presumably say, instead of 'the answer is in the content of the following lecture', that 'the answer is in the fact that the lecture is communicated'.

The Consciousness-Brain Relationship Is Like No Other.

We have got a bit too far ahead. The explanation for the problem will manifest itself in the lecture below, so let us reverse the order of the discussion and begin by thinking that a mind or consciousness of each person really exists in the world. If we think so, the fact that a mind or consciousness is produced by a kind of physical thing would certainly be puzzling. Why does a brain, a particular kind of physical thing, create a mind or consciousness, something of a quite different kind? Or rather, the question should be put as follows. It may be said that a physical thing produces or creates consciousness. But it is not important how we describe the relation. The question is: What is it for there to be such a relationship in the first place?

Even if it is true that the brain produces consciousness, the job the brain performs can never be seen, however closely the brain is observed. With regard to all other things in the world, if we make close observations of them, the things they contain or the jobs they perform will gradually become clear. However, there is no such ordinary connection at all between the brain and its job. The job the stomach performs can be found if we closely

observe it, but that performed by the brain could never be found by such means. In order to see what the brain does, I have to see the world, rather than observe or attend to the brain itself. The brown curtain I see in front of my eyes, the humming of the air conditioner I hear with my ears, and the taste of the gum I am chewing are the job the brain is performing. Nothing of a similar kind to this is happening anywhere else in the world. The relation between the brain and consciousness is similar to no other relations.

It is not similar to anything! That is true. And what is not similar to anything cannot be explained. For it could not be said that a certain sort of thing generally happens in a certain sort of case, and that the case in question is an instance of it. It seems that humans have a strong distaste for such a situation, tending to seek similarities of the case at issue with some other case. Humans try to gain comfort by taking the case to be a mere instance of relations commonly found in the world, e.g. by taking the relationship between a brain state and consciousness to be something like the relationship between electric discharge and lightning.

Of course, the relationship between the brain and consciousness is not similar to anything else, so it must be that the above kind of explanation is nonsense. It does not question what should be questioned, and is, in that sense, literally out of the question.

The problem might be regarded as similar to, for example, a hypothetical situation where we observe stones and notice that some of them generate heat and shoot out fire when moved. One may take ‘stones’ as a metaphor for organs, interpreting the situation as meaning that the brain reacts in quite a different way from other organs, or one may understand ‘stones’ in a broader sense as a metaphor for physical things in general. However closely we observe a stone that starts to burn when moved – which corresponds to the brain – we do not find it particularly different from other ordinary stones. Let us suppose that the difference is only in that it is a little whitish. The mind-body problem is commonly understood at this level of puzzlement. But that is entirely wrong. The real problem consists in the fact that in the case of the mind-body problem, unlike the case of the stone, what corresponds to the stone’s being moved and firing can never be observed. Therefore, it is already incorrect to formulate the problem as that of why the brain alone produces consciousness. For where are we to find ‘consciousness produced by the brain’?

The point is this: Nothing that can be publicly observed happens. That is, nothing analogous to the generation of heat or fire is happening; there is no mysterious flame observable by everyone. Indeed, it is unknown what is occurring. That is the problem.

In that sense, the more accurate metaphor is as follows. Suppose everyone had a box with something in it: we call it an ‘eetle’. But no one can see another person’s eetle. That is to say, an eetle is private. However, an eetle is linked with an external causal relation that can be seen by everyone. When someone touches the box (which is called an ‘eetle box’),

for example, the eetle goes into the 'ceetle' state. At the same time, part of the eetle box gets dented, and an invisible change also occurs in the box's microphysical state. When the eetle box is hit hard, the eetle goes into the 'deetle' state. At the same time, the box itself shrinks, and a different kind of change in its microphysical state also occurs. As a change observable only by the owner of the box occurs, a change of state observable from the outside occurs at the same time. Moreover, there are two kinds of changes of state observable from the outside, i.e. change of form that is visible for anyone, such as getting dented and shrinking, and change of state, which is microphysical, found only by specialists.

Of course, the owner himself can directly perceive 'ceetleness' or 'deetleness', and if the eetle turns into a ceetle or deetle without any causation observable from the outside, he can recognize the change. Each of the two sorts of changes observable from the outside can also occur independently of other changes. For example, even when the eetle does not become a ceetle and the eetle box is not dented, a microscopic change can occur. Even when the eetle does not become a ceetle and there is no microscopic change, the eetle box can get dented. Changes are also possible in all other combinations.

Nevertheless, the understanding of the meaning of 'eetle', at least at first, enters being related to the link between the eetle box's being touched and its becoming dented, and to the link between the box's being hit and its shrinking. That is particularly clear when a child is first taught by an adult the meanings of the 'ceetle' and 'deetle' states, and thereby the meaning of 'eetle'. Neither a microphysical change nor the state of the eetle itself seen by the owner is involved here. For when an adult teaches a child these words, the adult knows neither the microphysical state of the eetle box nor the state of the eetle itself that the child sees. Regardless of how the eetle appears to the child, or what the microphysical state of the child's eetle box might be, the eetle's state that the child sees when the box is touched and is dented is the 'ceetle' state, and the eetle's state that the child sees when the box is hit and shrinks is the 'deetle' state. There is simply no other criterion.

However, naturally, there are putative refutations from the other two sides.

The first refutation comes when the child starts to grow up. The child becomes able to recognize the change if the eetle has changed into a ceetle, for some reason, without the eetle box being touched or getting dented. The child becomes able to recognize the change if the eetle has turned into a deetle, for some reason, without the box being hit or having shrunk.

It is now necessary to bring the 'eetle' metaphor back to the actual situation prior to the metaphor. For if the supposition is that one cannot see another's eetle but can, when one's eetle comes to be in the 'ceetle' or 'deetle' state, directly 'see' the change of state in one's own box, it would be possible to describe what is seen in the box by comparing it to something that everyone can see in the external world. One might, for example, say, 'An eetle

is like round bread'. In reality, even that is impossible. I cannot describe the 'pain', 'sourness', 'anxiety' and 'melancholy' that I feel by comparing them to external things perceivable by everyone in the same way. For example, 'sourness' is linked with the common world shared by everyone only through my eating a pickled plum or a Watson pomelo and my face looking as if I had something sour, just as the eefle box is touched and gets dented. I cannot explain the sourness itself that I feel based on its similarity to something in the external world that is perceivable by everyone in the same way.

In spite of this, the first criticism is nevertheless possible. Suppose that even though I had not eaten anything sour, my mouth suddenly became full of a sour taste (and my face looks as if I have just eaten something spicy). I would be able to directly recognize how it feels, even though I could not explain this feeling to others. (The best I could do would be to say that it feel as if I ate a pickled plum or a Watson pomelo, demonstrating what the normal facial expression would be, although it is neither that I ate such a fruit nor that I have that normal facial expression.)

This first counterattack – i.e. the independence of the cognition of sensation – is crucially important. For it is the possibility of this counterattack that makes pain, sourness, anxiety, melancholy, etc. 'what can be felt', that is, 'sensations' and 'emotions'.

For the following lectures, I would like to introduce new terms in advance. In the case of sourness, what is *regarded* as being felt when eating a pickled plum or a Watson pomelo is the 'primary intension' of sourness. The feeling itself at the stage where, after the first counterattack, it has become possible to sense sourness, for some reason, without eating anything sour is the 'pre-primary intension' of sourness. It is important that the pre-primary intension is reached via the first counterattack. The 'secondary intension', which is introduced by the second counterattack, will be explained later.

One might think that the first counterattack has not succeeded. The issue might be clear if we consider the case of right and left, which, unlike pain, sourness or melancholy, does not involve privacy. A child who has just learned the distinction between right and left normally cannot use that distinction from another person's point of view. There is the stage where a child can easily follow the order 'Raise your left hand', but cannot easily follow the order 'Touch Daddy's left hand'. That is, there is the stage at which a child only grasps the pre-primary intension, although it has learned the word through the primary intension. Naturally, there is the same stage for the word 'I'. Has a child at this stage, then, grasped the concepts of right and left? Yes, it has, for it can properly use these words. However, one who *teaches* the concepts of right and left has to be at a more advanced stage of understanding than the child in order to teach the child the concept in the first place. So, too does one who *understands* the child's utterance (such as 'My left leg hurts'). Since the child can properly use the words 'right'

and 'left' by saying, for example, 'My left leg hurts', they could be said to have acquired the concepts of right and left. But that is because there is supplementation by the interpreter. The concepts of right and left would have no currency in a world in which everyone grasped them in the same way as a child.

The case of 'right' and 'left' and that of 'sourness', however, differ as follows: whereas one's assertion 'This direction is to my left' would not be accepted if the direction is, objectively, to one's right, the assertion 'This tastes sour to me' would sometimes have to be accepted even if what one has in the mouth is, objectively, salt or sugar. It would then be the case that 'sourness' exists. That is how the first counterattack succeeds.

The second possible criticism of the argument is related to physical states. It could be discovered, for example, that the real essence of water is H₂O, and the familiar wateriness is a mere contingent property of that real essence, or that the real essence of heat is molecular motion, and the hotness we feel is a mere contingent property of that real essence. (It would then be possible that something apparently indistinguishable from water is not in fact water because it is not H₂O, or conversely, that something that does not appear as water but rather as, for example, iron is in fact water because it is H₂O.) Similarly, it could be discovered that the real essence of ceetleness or deetleness is the microphysical state of a box, and ceetleness or deetleness is what it contingently accompanies.

Such discoveries of real essences must, rather than could, be possible, because of our most basic impulse to understand the world as complete *without us*. This might be regarded as a necessary consequence of the ideal of cognition. We are beings that cognize the world, and the way the world is must be *ipso facto* separated from our own cognition of it. It must be that even if there were no other organism that felt heat like us, heat itself would still exist irrespective of such a contingent state of affairs. It must be that heat *happens to* feel to us the way it does. Similarly, it must be that whether or not sourness feels the way it does to us, the essence of sourness really exists in itself, irrespective of such a contingent state of affairs. It must be that sourness *happens to* feel to us the way it does.

Conversely, there is no possibility of such an inquiry as to a chair, bookcase, baseball, the national constitution, a bachelor and so on. So there is no possibility of discovering their essences. Because they are artificial, what they are, or their essences, are in our hands from the outset. We can say, to use technical terms, that there is no possibility of separation between a priority and necessity. Nevertheless, it is not impossible that, as a result of investigating the biological nature of the human behaviour of sitting, certain kinds of chairs are discovered not to be chairs, and things that we thought were not chairs – e.g. straps in a train – are discovered to be chairs. Since humans are not wholly autonomous (self-controlling) organisms, being part of nature, that sort of inquiry is always possible. There could always

be the possibility for discovery in the case of ideologies such as feminism and libertarianism; there could be a true essence of feminism or libertarianism which feminists or libertarians may not know.

This also means that it is possible for a so-called analytic truth such as ‘Bachelors are unmarried’ to be falsified. This is not for the insignificant reason that the usage of the word ‘bachelor’ can change due to some contingent circumstances, but for the more profound reason that the natural essence of the institutional concept ‘bachelor’ can be discovered – i.e. that institutional facts such as someone’s being a bachelor can also have a natural basis.

Let us revisit the main argument. Once the physical states criticism succeeds, the following becomes possible: If one, without eating anything sour, feels as though the mouth is filled with a sour taste, examination of one’s nerves and brain may describe the feeling as an ‘illusion’. Or, to use another example, however intense a pain is, if it is due to a disorder of the sensory areas of the cerebral cortex, there being no normal cause at the nerve endings or no excitation of C- or A δ - fibres, it could be described as ‘the illusion of pain’, and that therefore there is in fact no ‘pain’. The converse would also be possible; if there was a normal cause at a nerve ending and an excitation of C- and A δ - fibres, someone who felt nothing, perhaps because of a disorder of the cerebral cortex, could be said in fact to have ‘pain’.

This all serves to highlight the importance of the physical states criticism. It is the very possibility of this criticism that locates ‘the mental’ such as pain and sourness in the communal world as ‘objective’ entities. A signal from a nerve terminal is sent through nerve fibres to the spinal dorsal horn, and then to the sensory areas of the cerebral cortex and to the thalamus. It is crucially important that this is a public process located in objective space, and that it has objectivity such that it is in principle observable by anyone. ‘Treatment’ of pain becomes possible through the mediation of this objectivity!

One might wonder if the second counterattack can ever succeed. It might seem that, regardless of the state of some fibre or other, there is pain if pain is felt and there is no pain if no pain is felt. If so, the feeling of pain itself can never be disregarded. That is indeed true. The implication here is that the first counterattack (i.e. the independence of the cognition of sensation) can be repeated against the second counterattack (i.e. the physical states argument).

Physicists identify heat as molecular motion. So even when something is felt to be hot, heat is taken not to exist if there is no molecular motion. Moreover, even when nothing is felt to be hot, heat is taken to exist if there is molecular motion. Nevertheless, there still has to remain the sensation of hotness itself, which is graspable independently of heat (i.e. molecular motion). For if not, there would be no such thing as ‘what has been identified with molecular motion’. Once the meaning of ‘heat’ is taken over by that of ‘molecular motion’, it

can no longer be possible to discover that heat (i.e. molecular motion) is in fact not molecular motion. Yet it remains possible to discover that the sensation of hotness is not caused by heat (i.e. molecular motion). The same is true of pain, sourness, anxiety and melancholy. Otherwise, there would be no such thing as 'what has been identified with excitation of C-fibres'. What is identified with something, whatever its identity, must permanently remain as that which is graspable independently of that identity.

In fact, it is not only that it permanently remains so, but that the essence of mental states such as pain, sourness, anxiety and melancholy is the way they appear to the person who feels them. Yet the real essence of the mental can only be distinct from their appearance. What must be an appearance by definition, i.e. what cannot be a real essence by definition, would not change at all by being defined, even if its real essence were discovered. The fact that there can exist that which would not change and the presupposition that there exist conscious subjects – that is, us – who are constituted by such resistance to change are two sides of the same coin. The assumption that there exists an objective world independent of us, which we investigate in an attempt to discover its real essence, may be a firm one. Nevertheless, the presupposition that phenomena appearing to us can be captured independently of the grasping of their real essence or identity constitutes a presupposition inseparable from the assumption that there exists an objective world independent of us.

Is this not a very curious situation? However much we may stress that 'the essence of the mental such as pain, sourness, anxiety and melancholy is the way it appears to the subject of that feeling' (and so that real essence or identity is not essential), none of us knows the way it 'appears' to others. Indeed, we do not even know if an 'appearance' exists for others. However much we may emphasize the fact that sourness, independently of its real essence, feels to us the way it does, none of us knows the way it feels to 'us'. Or rather, there might not be a common way sourness feels to us all. We simply do not have the means to find out whether or not it does.

A change in an eettle is fundamentally different from a change in a stone. We might decide that ceetleness is in fact some microphysical state, but we do not know what *general* 'ceetleness' is. So we would not know *what* is being said to be the microphysical state. This fact is presupposed by the rivalry between the three conflicting accounts – the rivalry between causation manifested by external behaviour, the 'familiar' sensation which immediately 'appears' to us, and the internal physical state obtaining in our physical body. Therefore, there is no case similar to this rivalry. Nothing similar to it is happening in the world. Philosophical disputes arise concerning what would be similar to this case, which is not similar to anything. The examples of the relationship between lightning and electric discharge and that of the relation between water and H₂O do not offer good analogies because

a sensation, which is immediately felt only by a self, is said to correspond to lightning or water, which is observable by anyone. One of the pleasures of philosophy as an intellectual game is finding out a similarity that is not commonly noticed – a similarity like that between the actor Ryotaro Sugi and the singer Masahiko Kondo. Are they still famous?

An analogy with time

Let us try an analogy using consciousness and time. We can draw an analogy using the relationship between an immediately present sensation and a present event, or another person's cognition of the self's behaviour and the diachronic record of a present event, or a physical state inside the body and some atemporal truth.

A present event is not something that is immediately experienced *only* by a self. It is, however, immediately experienced only at that time. (Note, however, that the present and past differ from self and other in that between present and past there is a direct link, namely memory.) Then, memory and all other things conveying events which were present to the new present correspond to external behaviours linking self and other, and physical facts, which are unrelated to such intersubjective relationships or subjective cognition, correspond to the objective event order, which is unrelated to such temporal modalities as past, present and future.

I think that, compared to the others, the time analogy is the most appropriate one. Moreover, by adopting this analogy, it becomes manifest that the following *two* problems have been hidden.

The first problem is this: Could it not be that the past and future are present memory and present anticipation respectively, and that atemporal truths are truths that are believed to be atemporal at present? We could think that past, future, and atemporal truths themselves simply do not exist. I do not mean to claim this myself, but the fact that we could do so is very significant.

Our analogy, then, yields this thesis: external behaviour and physiochemical states are things known by observation – that is, they are all the objects of someone's perception. This is in a sense true. Eating a Watson pomelo and acting as if it tastes sour cannot be known unless someone observes it, and the state of the brain and nerves obtaining then can only be known if someone observes them. To use the 'eetle' metaphor, everything is inside an 'eetle box'. The essence of the 'eetle' metaphor, in fact, consists in that point. I have said as follows: if the supposition is that one cannot see another's eetle but can, when one's eetle comes to be in the 'ceetle' or 'deetle' state, directly 'see' the change of state in one's own box, it would be possible to describe what is seen in the box by comparing it to something that everyone can

see in the external world. However, since 'something that everyone can see in the external world' is also inside an 'eetle box', it is in fact not seen by everyone. *Seeing* an eetle box's being touched and becoming dented could be 'feetle', and *observing* a microphysical change could be 'geetle'. Just as I cannot describe the 'pain', 'sourness', 'anxiety' and 'melancholy' that I feel by comparing them to what can be experienced by everyone, so I cannot describe the 'house', 'sky', 'someone's face after eating something sour' and 'physical state of another person's brain' which I see by comparing them to what can be experienced by everyone. There is, in fact, nothing that can be experienced by everyone.

Simply put, the idea here is that everything exists inside the mind. (Moreover, this idea is now being based on the fact that no one knows what is in others' minds, or that no one even knows if they exist at all.) In a sense, is that not simply true? All the facts in the 'external world' have to be known by means of someone's perception. However, although we ordinarily say that we cannot feel others' 'pain', 'sourness', 'anxiety', 'melancholy' and so on, we do not say that we cannot see the 'house', 'sky', 'someone's face after eating something sour', or 'another's brain' which others can see. Why? What difference is there?

Let us consider an intermediate case – i.e. the case of colour. Although we cannot see the colours red and green that others see, we live with others distinguishing between red and green in a common way. When there are red and green books, the request, 'Please get me the red book', is properly understood regardless of how 'red' looks to the person who hears the request. It is sufficient insofar as red and green are correctly distinguished. We have been taught that the colour of a sunset, fire engine, blood, tomato, etc. is 'red', and the colour of tree leaves, grass, etc. 'green'. When we were taught these words, the adults teaching them could not know how the colours looked to us. The colours we saw were irrelevant. (This is precisely the same as the case of learning the word 'sour'.) If it is commonly acknowledged that a sunset, fire engine, blood, tomato, etc. have a similar property, that tree leaves, grass, etc. have a similar property, and that these two properties are of the same kind but not similar to each other, communication about 'red' and 'green' can be established. The dysfunction caused by red-green colour blindness is the inability to distinguish between the two properties.

In contrast, there is no dysfunction caused by red-green colour inversion. A red-green inverted person is one who sees red when looking at tree leaves and grass, and sees green when looking at a sunset, fire engine, blood and tomato. Is there anyone red-green inverted? We can never know if there is. No possible progress in the relevant sciences will provide the answer. That is obvious because progress in the relevant sciences is progress of the second counterattack we discussed above. If I connect a nerve fibre from another person's brain to my brain in an attempt to see the green colour seen by that person or to feel the sour taste felt

by him, what I see or feel is still my colour or taste rather than his. That is, it is by no means possible to know whether the colour or taste I experience is the same as what is experienced by the other person. Or rather, there is in principle no such sameness or difference in this realm. That is the true meaning of there existing another. We will discuss this in detail in the next lecture.

The tentative solution to the problem here is to go beyond the second counterattack (i.e. beyond the ‘physical states argument’) and return to the stage before the first counterattack (i.e. before the ‘independence of the cognition of sensation argument’). Although it is tentative, there is presumably no other solution.

Because a child is taught that the colour of a sunset, fire engine, blood, tomato, etc. is ‘red’, and that the colour of tree leaves, grass, etc. is ‘green’, and because how the colours look to the child is disregarded when the words are taught, the colour of a sunset, fire engine, blood, tomato, etc. *is* ‘red’, and the colour of tree leaves, grass, etc. *is* ‘green’. How the colours look to the child is irrelevant. Precisely the same is true of the word ‘sour’.

Of course, after we grow up, we could report that a sunset, fire engine, blood, tomato, etc. have come to look green, and tree leaves, grass, etc. have come to look red for some reason. If colour inversion occurs between the right and left eyes, we have the right to report it to a doctor. Such inversion is certainly possible. Nevertheless, there is *no* possibility for such inversion to happen to a child at the stage of learning a language. For that is the starting point of language acquisition. In that sense, there is no one red-green inverted. Qualia inversion is impossible.

In spite of that, it seems to us that qualia still exist – that there still is redness itself or sourness itself which we feel privately (even though it performs no function). Why? This is the central theme of our lecture series.

Now, for the sake of discussion, let us draw a distinction between immediately feeling something and discerning something from other things. Whilst to discern is functional, to feel immediately is substantial. That is, whilst to discern is perceptual, to feel immediately is sensory. In other words, to discern is a ‘*psychological* act’, and to feel immediately is a ‘*phenomenal* fact’. The latter, when its vivid quality is emphasized, is sometimes called ‘qualia’, or simply referred to as ‘experience’. It can also be said that to discern is an act of the ‘mind’, whereas to feel immediately is a fact of ‘consciousness’.

To use the first two of the above contrasts, a substantial, sensory element plays an essential role in the case of ‘pain’, ‘sourness’, ‘anxiety’ or ‘melancholy’. They are ‘felt’. On the other hand, in the case of a ‘house’, ‘sky’, ‘someone’s face after eating something sour’, or ‘another’s brain’ (which are ‘seen’), a functional, perceptual element is essential; a substantial, sensory element (though it should of course exist) does not play an essential role.

This is the difference between the two groups of cases. If I use the above distinction between the 'mind' and 'consciousness', the theme of this lecture series, then, is not 'Is the mind real?' but rather 'Is consciousness real?'.

The present argument started with an analogy relating to consciousness and time. We have analogized the sensation or consciousness immediately present to a self in relation to a present event; causation manifested by the self's behaviour in contrast to a diachronic record of the same event; and the physical state inside the body with atemporal truths (concerning which we disregard such temporal modalities as past, present and future). Then, if what we have said regarding linguistic expressions of what is immediately experienced *only by the self* is right, we should be able to, by using the analogy the other way, say the same regarding what is immediately experienced *only at the time*. As we have discussed, the colour of a sunset, fire engine, blood, tomato, etc. *is* 'red', and the colour of tree leaves, grass, etc. *is* 'green', whereas how the colours look to a child is irrelevant. Similarly, all understanding of meanings in present experience is essentially cross-temporal, and any feature peculiar to the present experience is irrelevant to this understanding. For example, in the constitution of the meaning of 'sourness', the presentness of 'I actually feel sour now' cannot play an essential role; only the content it shares with 'I felt sour' and 'I will feel sour' can have an essential role. This is, in a sense, a matter of course because, just as we have to learn language from another, so we have to learn it from memory.

When I said that the analogy with time was the most appropriate, I also held that 'by adopting this analogy, it becomes manifest that the following *two* problems have been hidden'. I stated the first problem as follows: 'Could it not be that the past and future are present memory and present anticipation respectively, and that atemporal truths are truths that are believed to be atemporal at present?' Through developing discussions on this problem, we have arrived at a distinction between the mind and consciousness. But it remains a complete mystery what 'consciousness', which a 'self' *should* have privately, is. It is our task to reveal the essence of this mystery.

Another, more important, analogy with time

Now, what is the second hidden problem? It stems from the double meaning of the concept 'present'. I have said that the *presentness* of 'I actually feel sour now' cannot play an essential role in the constitution of the meaning of 'sourness'. The problem is: *When* is this 'present'? What is meant by 'actually now'? Does it mean 'only at this present time'? Or does it mean 'at each time'?

The word 'present' has two meanings. In the first meaning, there is only one present.

The present is only actually here. It is, of course, true that there was a present time at every time in the past, and will be a present time at every time in the future, but they are not the real presents. The real present is only actually here now. Is this not, in a sense, utterly self-evident? However, in the other meaning, there is a present time at any time. If a subject with a reflective consciousness exists, the time at which she is conscious is a present time. So, there is no such thing as the sole real present. Each and any time is ‘the sole real present’. There is only a plurality of ‘onenesses’! Then, the reason why we say, ‘It is the twenty-first century now’, would be merely because we are in the twenty-first century. Our saying so would be precisely on a par with someone in the sixteenth century saying, ‘It is the sixteenth century now’, or someone in the twenty-fourth century saying, ‘It is the twenty-fourth century now’. In a sense, is that not also utterly self-evident?

It could be said that the latter meaning is not self-evident. One could hold a position as follows: A time at which a reflective subject exists, and is conscious, and conscious of the time at which she is, *is* a present time. But, one might contend, such a conscious act is *actually* performed only at this time, and if we take into consideration what is ‘*actually* performed’, it has to be that the sole real present is actually here. One might say that the element ‘actually’ is indispensable for dealing with a problem about the present. However, in fact, a person in the sixteenth century or a person in the twenty-fourth century could use the word ‘actually’ to say precisely the same thing. This would make it impossible for any number of words to express the fact that there is the sole real present here.

What we can clearly see here is the conflict between the linguistic world established through language and the prelinguistic world which cannot be expressed by language. This may sound surprising, but the source of the mystery of ‘consciousness’, in fact, lies in a conflict of that kind. If this problem is not involved, there is nothing particularly puzzling about consciousness. That is what I would like to illuminate in our lecture series.

Now, the ‘present’ has been introduced as a temporal analogy to a ‘self’. The idea was that the sensation or consciousness immediately present to a ‘self’ could be analogized with a ‘present’ event. The problem which has arisen concerning a ‘present’ time also arises concerning a ‘self’ in that sense. That is, a ‘self’ also has two meanings. In one meaning, there is only one self. The self is only actually here. A ‘self’ in this sense would be better expressed as ‘I’. Of course, others would all insist that they are ‘I’s’, but they are not the real I. Another might say, “‘I’ refers to me”, but that ‘me’ is merely another’s ‘me’. The real I is only this I who am actually here. As in the case of the ‘present’, that is utterly self-evident. However, clearly the manner of speaking here is in more conflict with the way the linguistic world we inhabit is than in the case of the ‘present’. We will consider the reason for this later.

In the other meaning, anyone is a self. If someone with a reflective function is reflectively

conscious of herself, she, being reflectively conscious of herself, *is* a self. So, of course, there is no such thing as the sole special self. For each subject there is 'the sole special self'. There exists a plurality of 'onenesses'! Then, the only legitimate answer to the question, 'Why am I Hitoshi Nagai?' asked by me would be, 'Because the question is asked by Hitoshi Nagai'. This is just as the only legitimate answer to the question, 'Why is it the twenty-first century now?' asked by us would be, 'Because the question is asked in the twenty-first century'. The question, 'Why is it the sixteenth century now?' asked in the sixteenth century would be answered, 'Because the question is asked in the sixteenth century', and the question, 'Why is it the twenty-fourth century now?' asked in the twenty-fourth century would be answered, 'Because the question is asked in the twenty-fourth century'. Since these pairs of questions and answers are all precisely on a par with one another, there exists no special actual 'present'. A 'self' could be understood analogously to this, and a 'self' in that sense would be better expressed as 'one reflectively conscious of oneself', as opposed to 'I'. Thus, a 'self' has a double meaning; it means either 'I' or 'one reflectively conscious of oneself'.

Here, again, it is possible to insist as follows. It has been said that if someone with a reflective function is reflectively conscious of herself, she *is* a self. However, I might contend, it is only I who am *actually* reflectively conscious, and if the *actual* reflective consciousness enters into consideration, it has to be that the sole real self is actually here. The claim would be that the element 'actually' is indispensable for dealing with a problem about the self. However, since anyone could use the word 'actually' to say precisely the same thing, it would be impossible, after all, to express the fact that there is the sole actual self here.

What manifests itself clearly here is, again, the conflict between the linguistic world established through language and the prelinguistic world which cannot be expressed by language. In the current case of 'I', it seems more difficult to express the conflict by language than in the case of the 'present' or 'now'. The reason is simple; it is because I, at any time, can talk with others, whereas we, in the present, cannot talk with people in the past or future. (Each of us has a powerful means of cross-temporal communication within an individual, namely memory, which is precisely what enables establishment of a self. But it still does not make it possible to 'talk *with*' anyone.) So no one will object to our assertion that the true present is only here. In that sense, this assertion could gain public approval by everyone. (In another sense, however, that only means a plain fact that the assertion is not objected to for now.) In contrast, if I say that the true self is only me, all other people could object to me on the spot. That is, in principle, my assertion could not gain approval by anyone. (In another sense, however, that only means an obvious fact that the assertion is not approved of by anyone outside of me.) The conflict between the linguistic world and prelinguistic world is inexpressible by language. It seems that that is true at a higher level in the case of 'I' than in

the case of ‘now’.

However, if language is regarded as what is used to communicate something to oneself – that is, as what is essentially used to write a diary – rather than as what is used to talk with others, the situation is precisely reversed. If I, at present, write in my diary (under a certain date), ‘The true present is only here’, I, reading the diary at any time, could not agree to that. The assertion, in that sense, could not gain approval by anyone. On the other hand, if I write in my diary, ‘I am the only true self’, I at any time could agree to that. That is, no one would, in principle, object to the assertion.

Is language essentially for talking with others or for communicating something to oneself cross-temporally? This question only indirectly bears on the problem at issue, but I would like to discuss it briefly because it *per se* is significant. Since language, when introduced, is taught by others, communication with others is indispensable. After that stage, however, the situation is reversed, such that the aspect of language as the means to communicate something to oneself becomes indispensable. That aspect, then, becomes independent. This reversal is crucial, and it is the outcome of this reversal and independence that someone, as we have discussed, becomes able to ‘think’, internally, that red things *have come* to look green, or that sour things *have come* to taste bitter.

Wittgenstein has discussed whether or not a ‘private language’ is possible, concluding that it is not. But his argument is clearly wrong. Language is impossible unless a private language is possible. Language becomes complete when the possibility of a private language *turns into* what is indispensable for language. However, an attempt to speak of that fact in the ordinary public language must fail, since we cannot speak of it unless we follow the workings of the meaning of the public language. Hence, we cannot speak of what we intend to speak of – that is, whatever we intend to say, something ‘correct’ has to come out. This problem, in one aspect, is isomorphic to the problem that would arise with the ‘present’ and ‘self’, but to avoid more complication, let us return to the problem with which we started.

The second hidden problem was that the ‘present’ has a double meaning. Now, what should be noted concerning this problem is that it has to do neither with a problem of whether there really was the past or there will really be the future, nor with a problem of whether there really is a self or self-consciousness for others. These problems are rather related to the first problem – the problem raised by the thought that the past and future exist only within the present. Failure to understand the difference here will spoil all the effort. The past and future may really exist. Other selves and self-consciousness also may. What is crucial is that there is still a problem independent of that. Understanding this, one can suspect that the first problem is, nevertheless, only derivative from the second problem. This bears on the central theme of our lectures, or rather, is the very theme.

Who is a 'self'?: 'I' vs. 'one reflectively conscious of oneself'

We are now returning to the problem with which we started. Our discussions began by refusing to understand the mystery of consciousness in terms of the analogy with a mysterious stone which, when moved, generates heat and shoots out fire. In the mind-body problem, unlike in the case of the mysterious stone, what corresponds to the stone's being moved and shooting fire is not publicly observable. Hence, as we have said, it is already incorrect to formulate the problem as that of why the brain alone produces consciousness. Nothing that can be seen by everyone, such as the generation of heat or fire, is occurring. What is puzzling is that we do not know what is occurring. That is why we adopted the analogy with an 'eetle'. Nobody can see others' 'eetles'. But a 'ceetle' or 'deetle' belonging to oneself is directly perceivable, so one could recognize the change if the eetle, for some reason, has changed into a ceetle or deetle without any external causal relation. That was the supposition.

Who, then, is 'oneself'? It is easy to see that there is the double meaning we have been discussing. Is it 'one reflectively conscious of oneself'? Or is it 'I'? Suppose that it is 'one reflectively conscious of oneself'. Then I may acknowledge that 'one reflectively conscious of oneself' could recognize the change if the eetle, for some reason, has changed into a ceetle or deetle. Nevertheless, I would not know what 'ceetleness' or 'deetleness' itself is like. Suppose, on the other hand, that 'oneself' means 'I'. Then I would know what 'ceetleness' or 'deetleness' itself is like by directly experiencing it. But since the authority by which one is taken to recognize a change of the eetle into a ceetle or deetle originates from one's being 'one reflectively conscious of oneself', the ceetleness or deetleness I actually feel is, in fact, irrelevant to this system of authority.

The following objection might arise: 'One reflectively conscious of oneself' should be directly experiencing the consciousness of oneself, and that is precisely the reason why the 'authority' is granted. This is a reiteration of the problem we have discussed before. It should certainly be true that the sixteenth century is the 'present time' during the sixteenth century, and the twenty-fourth century is the 'present time' in the twenty-fourth century. Nevertheless, they are not the actual present. Both in the case of the 'present' and in the case of the 'self', the gap between the actual and the possible will never be obliterated. What is more, the gap between actuality and possibility *per se* is turned into a possibility, resulting in the *concept* 'actuality', which is not actuality itself.

The same thing could be said from the opposite direction. It could be said as follows: The meaning of 'I' is no more than 'one who directly experiences the consciousness of oneself', and that is the reason why 'I' am given the authority to privately recognize a change of the eetle. This is analogous to saying that the twenty-first century is the 'present' for us living in

the twenty-first century, and that that is exactly on a par with the sixteenth century being the present for those who are in the sixteenth century, and the twenty-fourth century being the present for those who are in the twenty-fourth century.

Here lies the elusiveness of ‘consciousness’, which is something a ‘self’ alone can directly experience. ‘Consciousness’ eludes any attempt to grasp it conceptually, endlessly escaping into the direction of ‘I’, which is ungraspable. It eludes any attempt to grasp it immediately, escaping into the direction of ‘one reflectively conscious of oneself’, which is already grasped. Furthermore, the same thing happens again both within the grasping of ‘I’ and within the grasping of ‘one reflectively conscious of oneself’, complicating the problem even more. However, this is precisely the way ‘consciousness’ exists. It is something whose way of being is fundamentally different from things that are real in the ordinary sense, i.e. things that are real such that instances of a concept appear on the same plane.

Simply put, I only know the red I see, the sourness I taste and the melancholy I feel. That is, I only know the consciousness I have. This directness is of a different kind from the directness *each one*, whoever they may be, has when they are said to directly know something. For I do not know the directness *each one* has. But is this true? Are these two kinds of directness really different in kind? No. When they are communicated by language, they become the same in kind. Indeed, they are the same in kind when you read this. (I have pointed out that in the case of the ‘present’, the same thing happens with a diary.)

Consciousness is what language betrays at its outset, and yet, at the same time, is what is created by that very betrayal. This must not be taken as pretentious rhetoric. This is where the whole problem lies.

Humans and animals, existing in the world objectively, each have a mysterious thing called consciousness. Strangely enough, it is produced by a physical thing called a brain. What is the relation between those two things? I do not mean to deny such formulation of the problem. Neither do I mean to say that the grasp of the world presupposed in such formulation is wrong. It *is* correct. However, the task of philosophy is to elucidate how we have actually reached this state of understanding. If we pass over this task, everything is merely a castle in the air.

Discussion

Question: If there were aliens engaged in human-like activities, who had achieved prosperity with science and technology more advanced than ours, could it be possible that they entirely lacked consciousness?

Answer: That question is far more complex than it seems. In the sense relevant to the

question, it is never possible for us to know whether they really have consciousness. One problem is the meaning of 'never' here. Why 'never'? If we can 'never' know, is it still either that they *in fact* have consciousness or that they *in fact* lack consciousness? That is also an immense philosophical problem. Since the case concerns aliens, there is additionally the problem of interpretation of language (or interpretation of what seems to be language). They might not only behave as if they have consciousness, but use a word that could be appropriately translated as 'consciousness'. Then, one view would be that they are conscious if they use such a word, and if they do not, they are not conscious. This is another big problem. One question here would be this: is it not possible that even in the case where they do not use such a word, they *in fact* have consciousness, and that even in the case where they do use such a word, they *in fact* lack consciousness?

One answer to this question is that this sort of situation is possible only in the case of oneself. Even if I am incapable of any outward expression, it is possible that I am actually thinking or feeling something. One for whom such a thing is possible is me; that is how I distinguish myself from others. Since none of the aliens is me, that sort of situation is impossible for any of them. Reaching this point rekindles the problem whether 'I' should be construed as 'one reflectively conscious of oneself'. Indeed, interpretation of the second person's language fundamentally presupposes the operation of regarding the second person as another 'I', i.e. as 'one reflectively conscious of oneself'. By acknowledging that what I have just said of myself is possible for the second person, I approve her as another 'I'. Conversely, I am also reducing my expression of 'I' to a level which enables the mutual interpretation of the word 'I'.

Since all of the above processes involve an immense philosophical problem, it would be extremely challenging to answer the alien question.

Question: Supposing a roboticist was building a robot, does it follow from what you have just said that it would be impossible for the robot to become conscious?

Answer: Philosophy is very difficult. One and the same philosophical claim has to incorporate what verbally appear as two aspects – that is, the aspect from which the answer is, 'It would be impossible', and that from which the answer is, 'It would be possible'. To begin with, the answer is simply, 'It would be impossible'. For consciousness is unlike a thing in a box in that it is not something that might or might not happen to exist. It is generally too naïve to think that we can readily handle 'consciousness' by treating it as a 'thing' that might or might not exist in an ordinary sense. Nevertheless, it is important to understand why such a naïve view is possible. Unless it becomes possible to suppose that the robot could (or could not) be conscious by chance, our concept of 'consciousness' is not complete. In that sense, the answer has to be, 'It is possible'. Consciousness eventually has

to be regarded as an existing ‘thing’. I would like to start considering the mechanism through which that happens.

Question: In the lecture, you have explained as follows: Since ‘something that everyone can see in the external world’ is also inside an ‘eetle box’, ‘seeing’ an eetle box’s being touched and becoming dented could be ‘feetle’, and ‘observing’ a microphysical change could be ‘geetle’. If that is true, it does not follow that ‘eetles’ belong to all people, or are plural things of the same kind co-existing. Rather, one ‘eetle box’ contains everything else – including other eetles – inside.

Answer: When you say ‘one “eetle box”’, do you mean an arbitrary one? Or do you mean a particular one (that is perhaps yours)? That is the problem. Consciousness cannot be something that co-exists with other plural instances of the same kind. It naturally has a structure of being singular in the first place and embracing everything else. Yet it cannot completely embrace other consciousnesses, because all the consciousnesses which are not completely embraced are also in such a way that they embrace everything else. However, if it is said that consciousnesses embrace each other, they, again, become appositional. This is not a characteristic that consciousness contingently has, but is its nature. Or, more precisely, the structure here is consciousness. So the idea of the so-called ‘privacy of consciousness’ is insufficient. For one thing, it is not the case that consciousness contingently has the characteristic of being private, but rather, the aspect of the world called ‘being private’ is identical with consciousness. For another thing, it is not the case that privacy co-exists with other privacies, but, conversely, the impossibility for it to thus co-exist with other instances of privacy (when it is forced to co-exist with other instances of privacy) generates co-called privacy’.

Question: Then, is it appropriate to use an analogy of there being one reversed tin at the centre? Genpei Akasegawa turned a food tin inside out and created a ‘Tinned Universe’, which has the whole universe ‘inside’. Am I analogous to such a reversed tin existing in a world where there are many other normal tins?

Answer: I didn’t know Akasegawa’s work you mentioned, but a tin reversed inside out sounds very heart-warming. At the centre is a reversed tin with the universe inside. Yet other tins in that universe, too, are each at the centre, and are reversed. So that fact must have been anticipated in the beginning to be built in the meaning of ‘being reversed’ when it is said of the tin which is actually at the centre that it ‘is reversed’. And this explanation itself also applies equally to all other tins in the universe. This goes on *endlessly*.

However, that should not be visualized as a mutual embracing of co-existing things as if the picture was of the world of Leibnizian monads or of the world of the Avatamsaka Sutra. Rather, the primordial reversed tin at the centre is simply reversed, and the universe begins

from it. So, it is not at all a tin, nor is it reversed. The thought that the tin is the same as other tins except that it is reversed is only invented afterwards. Indeed, it could not be more different from other tins. It comes to view itself as a tin when it finds some kinship with other tins, setting aside the fact that it is reversed. Moreover, it imposes on other tins this process of insight in the converse order, coming to view them as things of the same kind as itself except that they happen not to be reversed. It thus takes other tins as viewing it in the same way. The tins come to view one another that way, though inside one universe. The result is the picture of mutual embracing.

I will draw some diagrams in the next lecture. When you see them, please remember this discussion on tins.

(translated by Shogo SHIMIZU)

