

Free Will and the Divergence Problem

Takuo AOYAMA*, Shogo SHIMIZU** and Yuki YAMADA***

Abstract

This paper presents what the authors call the ‘divergence problem’ regarding choosing between different future possibilities. As is discussed in the first half, the central issue of the problem is the difficulty of temporally locating the ‘active cause’ on the modal divergent diagram. In the second half of this paper, we discuss the ‘second-person freedom’ which is, strictly, neither compatibilist negative freedom nor incompatibilist positive freedom. The divergence problem leads us to two hypothetical views (i.e. the view of single-line determination and that of one-off chance), and these views bring humans closer to the *afree* side – i.e. outside of the contrast between being free and being unfree. The *afree* side is greatly different from the ordinary human side. This paper tries to secure the second-person freedom as a substitute for the ordinary human freedom while preventing the divergence problem from arising.

0. Preliminaries

In the first half of this paper (from Sections 1 to 4), we will present what we call the ‘divergence problem’ and clarify the problem concerning cases where a choice is made between different possibilities – i.e. between different possible present times – *in the passage of time*. In the second half (from Sections 5 to 7), we will give a consideration of freedom in order to prevent the divergence problem from arising.

Words or phrases such as ‘freedom’ and ‘free will’ will not be used in the main argument of the first half. For its aim is to resolve the confusion which we can often find in past philosophical disputes on freedom – the confusion that arises from discussing the problems peculiar to the topic of freedom and other more general problems together.¹

* The Research Institute for Time Studies, Yamaguchi University

** The University of Tokyo Center for Philosophy

*** Faculty of Arts and Science, Kyushu University

¹ The problems peculiar to freedom are, for example, problems as to under what conditions it becomes possible to attribute responsibility to an agent. They are interesting, but they provide a discussion containing the following danger. ‘[E]ven if the possibility of responsibility attribution entails freedom, it would not follow that the latter entails

In the last section, we will sketch humans as *afree beings* – who are outside of the contrast between being free and being *unfree* (in a sense we will explain). The important point will be that even if freedom in the compatibilist sense (e.g. the capacity for moral responsibility) is secured, humans – and their brains that are often personified as the subjects of manipulation – can be regarded as *afree*.

1. The divergence problem

Various possibilities of history are often illustrated by a tree diagram – a tree diagram whose branches diverge from the past to the future. On any point on the diagram, there will be a single past history, whereas there will be many future histories. By making a decision, humans seem to choose only one actual branch from those many possibilities. But is this true? If a branch is chosen by human decision, where in a tree diagram is the moment of decision?

Reflection on everyday experience would incline us to say something of the following sort. ‘I went to the hospital because I had a fever of over thirty-nine degrees Celsius. The moment I saw the thermometer, I decided to go to the hospital, and immediately left home. At that very moment the history of going to the hospital was chosen.’ When we think this way, the moment of decision seems to be at the diverging point of the history of going to the hospital and that of not going to the hospital.

However, this way of thinking leads to difficulty. For both of the diverging histories include the divergence point. History A of ‘going to the hospital’ and History B of ‘not going to the hospital’ share their divergence point as their part (**Figure 1**).² So, it is impossible for a decision at the divergence point – call it Decision X – to make History A actually follow. Whatever kind of event a decision may be, all the events at the divergence point exist in *both* histories.

We can say, ‘The agent went to the hospital because of Decision X’, or, ‘The

the former (or at least it is not obvious that the latter entails the former). Therefore, there has in fact been a danger from the beginning that the extracted conditions of the possibility of responsibility attribution may *contain components unnecessary* for the conditions of the existence of freedom’ (Mino, 2008, p.172, our translation and italics).

² To make a crude linguistic distinction between ‘going to the hospital’ and ‘not going to the hospital’ is merely to put a label on each history out of epistemological interest. Each of History A and History B has details not mentioned on the label – such as the number of steps taken to walk to the hospital – and therefore is a complete individual history. So, the bifurcating histories in **Figure 1** may be taken either as two individual histories or as abbreviations of two *sets* of histories. The latter means abbreviations of infinitely many variations of the history of going to the hospital and that of not going to the hospital, each variation being an individual possible history, classified into two sets.

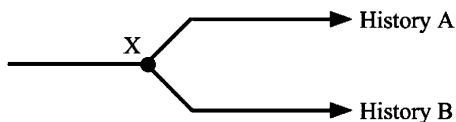


Figure 1 Decision X at the divergence point

agent would not have gone to the hospital without Decision X', if Decision X is later than the divergence point (**Figure 2**). Then Decision X can be regarded as a kind of ground related to going to the hospital. However, we would have to admit that Decision X is irrelevant to the choice of history. The 'ground' here is a ground in the sense that a counterfactual conditional analysis is applicable to it, but is not a ground for the choice of history. For Decision X occurs after History A is already chosen for some reason.

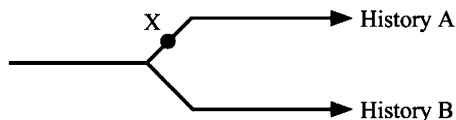


Figure 2 Decision X at a point later than the divergence point

As we will see in Section 3, a choice with no ground (for the choice of history) is mere one-off chance related to only one choice. But in the explanation we saw, the decision as the 'ground' could only be found in the history after the divergence point, and so the ground does not determine the choice of history. If Decision X is replaced with the event X in the natural world, a situation with exactly the same structure would be found in the case of a choice with no human intervention. How, then, do humans, *or nature*, choose a history? Let us call this the 'divergence problem'.

This problem concerns the temporal location of the 'active cause' of a choice that is made between future possibilities, but what prioritises the active cause among other causes? Of course, the authors of this paper endorse the general view of causation that one event can have more than one cause. (For example, one event can have causes that include a 'ground' to which a counterfactual conditional analysis is applicable.) However, what is important is that *unlike* such a general view of causation, which is compatible with determinism, the divergence problem concerns the question of how one of the possibilities is actualised as one point of time becomes *the actual present*. That is, what is called the 'active cause' in this paper bears not only on the untensed nomological relation between events, but on the tensed phenomenon of one of the possible present times *becoming actual* in an *indeterministic* world (cf. McTaggart, 1908; Horwich, 1987, ch. 2).

Since the form of the divergence problem is simple, the above presentation might

seem lacking in detail in the premises. But the apparent lack is deliberate. The divergence problem is raised as the catalyst that becomes active when connected to various hypothetical premises. From Sections 2 to 4, we will actually connect the divergence problem to different hypothetical premises to see what follows from each of them. For example, from a connection to one of the premises, a parallel to van Inwagen's famous argument for incompatibilism (e.g. van Inwagen, 1975) will be derived. But the simplicity of the divergence problem makes it possible to exclude the time asymmetry of the determination relation, which van Inwagen presupposes in defining 'T₀' in his argument (van Inwagen, 1975, p. 191), and to derive a more general, time-symmetric conclusion.

2. Nomological determinism

There is a leaf in the air. Which place will it fall onto? Let us tentatively assume that the laws of nature uniquely determine the destination of the withered leaf. 'Determination' here means derivability of the destination of the leaf from the combination of the past data – data concerning the leaf and its environment – and the laws of nature. (The limit of human computing or measuring capacity set aside). If there is data concerning the present moment, it can be fully worked out where the leaf will be ten seconds later, where it will be a hundred seconds later, or even where each of the atoms constituting it will be a thousand years later.

If the data concerning the world at the present moment is determined by the data concerning the world in the past (and the laws of nature), this tracing back in time will never end. Let us call S_1 the initial state of the world we inhabit, and suppose that another world whose initial state is S_2 is possible (supposing that the world has a beginning). To think of the divergence point between S_1 and S_2 would be to trace their histories back in time towards before the beginning of either world, and would be impossible by definition. The history following S_1 and the history following S_2 contain no divergence of possibilities inside them at all. All the events in each of the histories are determined by the initial state and the laws of nature. Choosing one of the initial states is equivalent to choosing the whole of the history it belongs to, and it is impossible to choose an initial state without choosing the whole history. That is, the choice here is not to be made at the birth of the world between the two initial states, but is to be made between the two wholes *quite independently of time*.

If the determination by the laws of nature is as strong as in the above account, there would be no room for the conception of diverging possible histories from the outset. For there would no longer be a reason to think of possibility in terms of a divergence towards the future. A possible history apart from our history, if any, would not be a history that diverged in the past, but would be an independent single-line history conceived of as utterly distinct from ours.

The fundamental laws of physics are said to be time-symmetric in several senses (e.g. Price, 1996; Horwich, 1987). This physical fact reinforces the above conclusion. If we consider the time symmetry of the determination by the laws, we can go on to say: if there is data concerning the present moment, it can be fully worked out where the leaf *was*, or where the atoms constituting it *were*, ten seconds ago, a hundred seconds ago, or a thousand years ago.

Just as data at one point of time determines data in the future, so data at one point of time determines data in the past. The implication is that data concerning an arbitrary point of time, not necessarily those concerning the initial state, would be sufficient to enable calculation of the whole of the history. Thus, if data at any moment of time is fixed (even if it is unknown to anyone, so long as it is fixed realistically), it follows that the data concerning the whole history is also fixed.

This is how determinism under time-symmetric laws encourages the elimination of the asymmetric, *determining-determined relation* – i.e. the elimination of the notion of ‘making someone do’ something or ‘being made to do’ something. (This determining-determined relation bears on our claims about the ‘free-unfree relation’ in Section 4 and about ‘unfree beings manipulated by something’ in Section 7.) A state at a particular point in time does not unilaterally determine a state at another point of time. It is only that two arbitrary states are nomologically related with each other such that one is calculable from the other.

3. Single-line determinism

Tautologically, probabilistic laws tell us about the probability of what will happen in the future. For example, they tell us that the probability of clear sky tomorrow is 80 percent. Yet they do not tell us how a *single* actuality is chosen from possibilities that are assigned different probabilities. To use a divergent diagram of time, probabilistic laws tell us various things about the possible situations immediately after the divergence point. Therefore, if there is decision-making based on probabilistic laws, its probability might be able to express the agent’s ‘inclination’ or ‘character’ (as, for example, ‘irritable’). However, probabilistic laws tell us nothing about the *genuinely one-off* choice of a course made at the divergence point – i.e. about which of the possible situations will become the singular actuality.

The following point is important: that a one-off choice happens by mere chance is compatible with the fact that the consequence of that choice has some coherence with things that preceded the choice. Let us look at Robert Kane’s example of a businesswoman (Kane, 1999, p. 225). In this example, there are two choices coherent with the agent’s preceding will (namely, the choice of attending an important meeting and that of helping the victim of an assault), and one of them is chosen by indeterministic chance. Then the choice itself is a mere one-off chance event, but the

consequence of the choice is coherent with the past and is *not* merely random. What is called ‘chance’ in this paper is this sort of chance consisting in a one-off choice, rather than the mere randomness of the consequences of a choice.

If the validity of probabilistic laws is not simply epistemological – that is, if it is not simply that the world that is really deterministic seems probabilistic to ignorant humans – such that its validity pertains to the way the world really is, there would naturally be a limited number of responses to the divergence problem. Put plainly, there would be only two responses.

The first one would be that a probabilistic event becomes the singular actuality by chance, and how one possibility is chosen is unexplainable. As we have just seen ‘chance’ here is nothing more or less than a one-off choice without a cause (or reason).³ The second response would be that all of the histories *qua* possibilities actually exist – that there is no choosing one from them. We will call the second response the ‘many-worlds view’ using this term in a broad sense, without relying on any particular theory such as particular interpretations of quantum mechanics. If the many-worlds view is right, the divergence problem would not arise. There would be no need to choose either of the histories at the divergence point.

One might say: it would still be true that there come to be two histories after the divergence point, so either of them would have to be chosen, otherwise, in order for both of them to be chosen, ‘now’ would have to split into two. The reason for this would be as follows. Up to the point of divergence, a single ‘now’ has moved in the single-line history, but after the divergence point, ‘now’ would have to separate into History A and History B. For this, ‘now’ would need to split at the divergence point.

The above thought involves confusion, because, in the first place, it is unreasonable to illustrate the many-worlds view by a tree diagram. If History A and History B both actually exist, they should be illustrated by two single lines, rather than by a tree diagram, as in **Figure 3**. The part of a tree diagram that appears as a single line is a part where the content of History A and that of History B overlap. That is, although their qualitative contents coincide, they differ as individual existences (cf. Lewis, 1981; Horwich, 1987, ch. 10).

On the many-worlds view, ‘I’ do not have both the possibilities of going and of

³ Although the concept of probabilistic causation is a significant one, it would not directly reject this response. The first major reason for this is provided by the plausible view that probabilistic causation concerns only general, rather than singular, causation, whereas the divergence problem concerns singular causation. That we use the expression ‘one-off’ corresponds to the fact that the divergence problem concerns singular causation. The second major reason is that even if the probability-raising in a certain case of singular causation is secured, there would remain unexplainable chance with regard to a one-off choice. Thus, concerning a case of singular causation, the occurrence of any consequence would not contradict any particular probability-raising (between 0 and 1, exclusive).

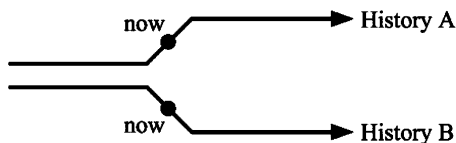


Figure 3 Two individual single-line histories

not going to the hospital. I am already in either of History A or History B, but since I am in the overlapping part, I only cannot tell in which of the histories I am. If I am in History A, I will go to the hospital, and if I am in History B, I will not go to the hospital. In the above picture, decision-making is like prediction. Thinking what to do next is similar to guessing what is true of the history one is in with respect to what one will do next. The reason why we humans ordinarily act naturally and feel that we are moving by our own will is that our prediction is often right. This is not a supernatural power or anything. Just as it is no wonder that we can correctly guess a past situation from the present situation, so it is no wonder that we can correctly guess a future situation from the present situation.

Now, is the current picture a kind of determinism? Here we need to draw a distinction between two kinds of determinism. One is determinism in which determinacy makes the history a single line, and the other is determinism in which the history's being a single line makes it determined. The former is nomological determinism, and the latter is what we will refer to as 'single-line determinism' when we use the term in a narrow sense.

To understand the former determinism, we should simply recall the commonly examined determinism. If the laws of physics determine what happens, branches of possibility would disappear, leaving only actuality. With actuality uniquely determined, the history would be a single line. On the other hand, the latter determinism is greatly different. If the latter determinism is right, there would be a single-line history at the outset. And it is because, and only because, that history is a single line that the history becomes determined. On this kind of determinism, whether or not we can know the determined content, what happens at every point of time is determined, without requiring laws.

After all, if either the nomological determinism or the single-line determinism in the narrow sense is right, all that remains is a single-line history. Therefore, we can collectively call the two kinds of determinism 'single-line determinism' by using the term in a broad sense. When we simply say 'single-line determinism' hereinbelow, the term will refer to that in the broad sense.

If we adopt the many-worlds view, all possibilities are actualised. All the possible histories each exist as a single-line history. So, in any possible history, single-line determinism is true. Even if a person, in any history, does not know what will happen,

what will happen is determined. What is interesting here is that the many-worlds view, which seems to entirely affirm the existence of possibility, becomes similar to the following view, which entirely denies the existence of possibility: ‘Possibilities are all products of our illusion, and there is simply only one history.’ For on the many-worlds view, all the meanings of such modal expressions as ‘necessary’, ‘contingent’ and ‘possible’ boil down to the features of the set of these single lines. (For example, what obtains in all single-line histories is ‘necessary’.) If we look at a separated individual single history (e.g. this actual history), modality would disappear.

4. Being unfree and being *afree*

Then, is the many-worlds view too absurd to pursue? Should we abandon it and think that the actual history alone exists? Would we then be able to properly understand possibility that possibly, though not actually, exists, or counterfactual possibility that *exists while not existing*?

In the field of possible worlds semantics, there are positions called possibilism and actualism. We will not repeat the well known problems concerning these positions, such as the problem of inevitable circularity concerning the definition of possible worlds in actualism (Loux, 1979) and that of trans-world identity in the sense of ‘overlap’ of worlds (Lewis, 1986, ch. 4). Instead, we will see in this section, both possibilism and actualism transform the concept of possibility into a non-ordinary one, and entail single-line determinism.

Let us consider a ‘possible’ world taken in its *ordinary* sense, that is, a world which could have been (rather than a possible world taken in the actualist sense). It should be what ‘does not exist’ actually but ‘exists’ possibly, and *if* it ‘exists’, it should not be a representation of a world but should be a world as totality. Most of the proponents of actualism who think that possible worlds are constructs inside the actual world cannot respond to this simple requirement. The reason is that this actual world is the sole existing world, and that possible worlds are non-existing things.

Possibilism overcomes this difficulty for actualism. According to possibilism, possible worlds are of course worlds, and they really exist. Nevertheless, recall that if all possible histories equally exist, single-line determinism would be true in each history after all. If ‘I’, in this world, do not go to the hospital, there is no possibility that *that I* go to the hospital. Every individual in the world could not have properties that *the* individual does not actually have.

Based on the above discussion, let us look at three ‘solutions’ to the divergence problem. The first and the most succinct ‘solution’ would be to eliminate the possibility of a choice in the first place. (As we said, this solution includes actualism.) If there is no such possibility, the divergence problem would of course not arise, and single-line determinism would hold for the world as it is in reality.

There are two other ‘solutions’. Let us recall the following argument concerning probabilistic laws. If probability pertains to the way the world really is, it would be that we should either accept one-off chance as something unexplainable or accept the many-worlds view (or possibilism). The former response to the divergence problem holds: a choice is certainly made, but there is no answer to how it is made. The latter response holds: all possible worlds are worlds in which single-line determinism is true and no choice is made. It is clear that these ‘solutions’ can be applied not only to probabilistic choices but to choices between possibilities in general.

The above ‘solutions’ can be summarised as follows: the understanding of possibility as an illusion (or as abstract constructs in the actual world) would eliminate possibility *qua* something that ‘exists while not existing’, and would make it something that simply ‘does not exist’. Acknowledgement of the many-worlds view would eliminate possibility *qua* something that ‘exists while not existing’, and would make it something that simply ‘exists’. We have seen that these ‘solutions’ (including actualism and possibilism) would both lead to *single-line determinism*. Furthermore, acknowledgement of unexplainable one-off chance would leave possibility *qua* that which ‘exists while not existing’ as something unexplainable. To acknowledge one-off chance means to grapple squarely with the divergence problem and to be completely pushed down by it.

In the world of single-line determinism, regardless of the truth of nomological determinism, the single-line history has no divergence of possibilities and no *active cause* at all. *Humans should then be called afree beings* rather than unfree beings. We model the contrast here on the contrast in ethics between being amoral and being immoral – i.e. between being outside morality and being opposed to morality. Being *afree* means being outside of the free-unfree relation *qua* the determining-determined relation. (We mentioned this time-asymmetric, determining-determined relation at the end of Section 2.) However, being *afree* does *not* mean being outside of the free-unfree relation in the compatibilist sense.

As we stated in Section 0, even if freedom in the compatibilist sense (or negative freedom, which will be discussed later) is secured, humans can be regarded as *afree*. For *afree* beings, it is never the case that the unilateral determination of one thing by another actualises one of the possible states of the former thing. Here there is no contrast between being free and being unfree that can be found in the determining-determined relation, and, in that sense, all events simply occur as they do. Many people would presumably see the shadow of Nietzsche’s or Spinoza’s philosophy in the picture of the *afree* world (e.g. Nietzsche, 1886, ch. 1, sec. 21).

On the other hand, if the answer to the divergence problem is sought in unexplainable one-off chance, it would be that any choice of a possibility is made by chance. The current case differs from that of single-line determinism in that possibilities (other than actuality) do exist, but if all choices of a possibility are governed by

one-off chance, all events simply occur as they do after all. We can say that in this case, too, *humans are afree beings* because there can be no contrast between being free and being unfree that can be found in the determining-determined relation.

5. Subtraction of raising from going up

The above considerations lead us to the concept ‘*afree*’. However, to understand the significance of this conclusion, we first have to examine the more familiar kinds of concepts of freedom in philosophy.

In the philosophical discussion of free will, a distinction is often made between two kinds of freedom. One is ‘negative’ freedom, which is freedom as absence of impediments or coercion by others. It is freedom to exercise the preceding will with no obstacle (e.g. freedom to claim what one desires to claim without political suppression). The other is ‘positive’ freedom, which is freedom as *spontaneous* choosing of possibilities. It is this freedom that faces the divergence problem.

The existence of negative freedom does not contradict nomological determination, and is compatible with both single-line determination and one-off chance. In this section, we will focus on positive freedom through interpreting Wittgenstein’s famous question in a new way, and we will also deepen our explanation of the determining-determined relation.

‘[W]hat is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?’ Wittgenstein posed this question in his *Philosophical Investigations* (Wittgenstein, 1953, Part I, §621). The standard answer by Wittgenstein scholars is: nothing – especially nothing like willing – is left over. But after the argument of this paper, we should respond in the following way: the result of Wittgenstein’s subtraction is less than zero, and the *indeterminacy* of the temporal position of the free subject is left over.

If we reverse Wittgenstein’s question and *subtract the raising of my arm from my arm’s going up*, the determinacy of the temporal position of the subject is left over (Aoyama, 2010, pp. 98–100). Through this the subject is captured as a temporal being. The subject becomes one of the ‘existing things’, that is, the subject becomes a kind of physical object. When I as the subject am thus temporally located, confrontation with the divergence problem would not allow that I choose neither in the realm of chance nor in the realm of necessity. The free subject as a non-temporal being vanishes, and I as a physical object become subject to single-line determination or to one-off chance. (To use the aforementioned expression, I become an *afree*, rather than unfree, being.)

We can link Wittgenstein’s question to the divergence problem by relating both of these to the so-called ‘problem of other minds’, especially in light of an argument by P. F. Strawson, which we will discuss later. The ‘problem of other minds’ in the

history of philosophy stems from the invisibility of the second person's mind. 'I' as the first-person subject only know my consciousness – the kind of consciousness dealt with in the philosophy of mind as 'qualia' or 'phenomena' – and do not know what another's consciousness is like. All the subjective experiences are literally concealed *subjectively*.

A point that is worth mentioning here is that, due to the intervention of the above problem of other minds, the visibility of positive freedom and that of phenomenal consciousness are both *inverse with respect to the first and other persons*. (Notice that the visibility of positive freedom is the visibility of its *non-existence*.) As we will see below, whereas the non-existence of the first-person's (or one's own) positive freedom is visible, the non-existence of another person's positive freedom is invisible. The invisibility of non-existence is, of course, quite different from the visibility of existence. Nevertheless, do we not *live* through the illogical replacement of the invisibility of non-existence with the visibility of existence? This is the question we would like to consider in what follows.

The existence of positive freedom is dubious especially in the first-person case. The more closely one observes phenomenal consciousness introspectively, the more it becomes clear that one cannot find the source of positive freedom over and above kinaesthetic sensations or anticipation of change of posture. As regards the introspective feeling of freedom, it remains possible that it is an illusion stemming from ignorance, and the feeling of freedom does not at all guarantee that the possibilities have been actually chosen (or even that there were possibilities other than actuality in the first place). Moreover, whichever phenomenal event we tentatively take to be the source of positive freedom, consideration of the temporal position of the event will lead us either to single-line determination or to one-off chance.

What is more troublesome is the problem of whether spontaneous choices of possibilities are really distinguishable from choices by mere chance. The concept of chance guarantees that possibilities exist, and that some of them are actualised with no restriction by past factors. It is extremely difficult to say what more is needed for 'spontaneity'. It would, of course, not suffice to say: spontaneity differs from chance in that spontaneity is actualisation of a specific intention. For there is adequate room for 'actualisation of a specific intention' also within the framework of negative freedom. (With regard to this point, the case of computer chess matches in Dennett, 2003 is illuminating.⁴)

⁴ Suppose that the chess programs A and B repeatedly play against each other. By using a pseudo-random number generator or a learning function, A and B can play in a different way each time. But the entire content of the repeated matches is nomologically determined. Now, if A wins more games than B, how should we explain the reason why A is a better player than B? Was it subject to nomological determination? Such an answer, as is pointed out in Dennett, 2003, pp. 77–83, is absurd. Perhaps A

After all, it is difficult to find the source of positive freedom inside the agent's first-person phenomenal consciousness. The same conclusion would be derived if another person's phenomenal consciousness was visible; the source of positive freedom would also not be found there. But another's phenomenal consciousness, especially its temporal position, is invisible. This invisibility, which stems from the problem of other minds, is significant when we consider the divergence problem. For it makes room for the idea that we posit a free subject behind another person's bodily behaviour precisely because of the invisibility of another's phenomenal consciousness. Let us explore the possibility of this idea in the next section by consulting Strawson's argument.

6. The second-person freedom

Strawson in his paper 'Freedom and Resentment' (Strawson, 1962) discusses *reactive attitudes*, such as anger, forgiveness and gratitude, from a viewpoint independent of the truth or falsity of determinism. Towards snow, goldfish, stationary and so on, we do not adopt reactive attitudes but adopt *objective attitudes*, such as treatment, management and avoidance. Even if determinism is true in some sense, Strawson says, it is impossible to only adopt objective attitudes towards every person.

The above argument is often understood as supporting a kind of compatibilism. But it is important that Strawson solely discusses moral practice, saying little about freedom (cf. footnote 1). We, the authors, admit that the above argument involves a deep insight. However, we doubt that the phenomenological fact that there is the practice of reactive attitudes supports the view that freedom is not threatened by determinism.

Instead of satisfying ourselves with the phenomenological fact that there is the practice of reactive attitudes today, let us discuss whether there is an element that induces humans to such practice – an element that might vanish when humans come to thoroughly believe in determinism. We think that the key to the element in question can be found in the very fact that *others are others*. That is, we think that what triggers a reactive attitude can be found in the invisibility of the second person's mind.

The above thought we propose is directly connected to the consideration of the divergence problem. For securing the invisible realm of second person's mind is very important not only for believing that there is the *active cause* of the action in the invisible realm (even if there is in fact no such cause), but for believing that the active cause is temporally located somewhere on the imaginary modal divergent diagram.

has a huge database of the standard moves or an effective system for scoring board positions. In that sense, A has the ability to realise a far more advanced intention than B.

In other words, the discussion below, in which we will deal with freedom on the basis of the fact that others are others, is a proposal for acquiring a substitute for positive freedom while preventing the divergence problem from arising.

A reactive attitude in practice should be, before it is a reaction to subjects in general (including oneself and unacquainted others), a reaction to a particular other person in front of oneself first and foremost. Kurt Vonnegut made a parody of Sartre's dictum and said, 'Inconvenience is other people, and inconvenience can be *Hell*' (Terkel, 2001, pp. 224–225). So far as negative emotion is concerned, another person can be the object of anger. It can be said that to feel anger, which is never felt towards snow, goldfish or stationary, towards another person means to give the highest level of ontological approval to the object, even though the emotion is negative. (This is presumably the reason why people often prefer being hated to being neglected.) The process here is that of approving of another person, an object with invisible interiority, which changes the movement of a body as a mere physical object into bodily behaviour by a *free* subject.

The same can be said of 'I'. When I as a first-person subject regard myself as free, or, more precisely, when I postdictively understand that a certain action by me *was* free, I am making myself another person, that is, I am viewing myself as one who is another person for another person. The psychological theories of self-perception and the similar behaviourist theories of self-interpretation have taught us that, in ascribing a psychological state to myself, data derivable by viewing myself as another person is often richer than those derivable from introspection. But we will not go into this here. What is important for this paper is that, conversely, the realm for the *invisibility* of the *non*-existence of free will can be secured by viewing myself as another person.⁵

So far as I observe my first-person experience, there is nothing invisible. The Freudian unconscious or the possibility of misrecognition of subjective emotion is conceptually demanded when I am observed and analysed under a particular theory, and such invisibility is distinct from the invisibility of the second person's mind. Of course, we are now considering the invisibility pertaining to the second person and the relation between that invisibility and reactive attitudes. Our hypothesis is this: 'I' regard another person as free and adopt a reactive attitude such as anger because there is, in addition to an observable action (e.g. a beating), an invisible realm in the second person, and a reason as a cause (e.g. ill will) is confabulated and posited in that invisible realm. If this is right, freedom, in that sense, is essentially second-person, and approval of the second person is itself connected with approval of freedom. As it

⁵ For example, it is significant for the creation of this realm that humans forget most of their phenomenal experiences. If humans remember all their phenomenal experiences with accurate temporal relations between them, the impossibility of temporally locating the active cause with regard to their own action would be more obvious.

were, *freedom is other people*. Let us further discuss this ‘second-person freedom’.

Another person has an invisible realm. This is different from saying that another person’s external behaviour is irrational. Rather, for a confabulation of a reason or cause it is necessary to see some regularity in another’s behaviour. If a person’s behaviour is utterly incoherent because of serious mental abnormality, it would be difficult to confabulate a reason or cause and posit it inside another person’s mind. On the other hand, if another’s external behaviour has a moderate degree of regularity, the regularity, understood as ‘personality’, ‘taste’, ‘belief’, etc., would contribute to the confabulation of a reason or cause. However, if the degree of regularity is *excessive* – such that, for example, the person’s next behaviour is completely predictable, as is the case with the behaviour of simple and non-random-number-based AI – the confabulation would be difficult.

The fact that verbalisation of the cause of an action is often unclear and is also a postdictive confabulation (e.g. Nisbett and Wilson, 1977; Johansson et al., 2005) does not mean that it is simply a result of cognitive error. Rather, if it was clear and was what identified the active cause of an action, the divergence problem, concerning the temporal position of that cause, would threaten people. As it were, the ‘cognitive error’ and the ignorance of the existence of the error make humans believe that the active cause is temporally located somewhere on the imaginary modal divergent diagram (even if there is no such cause).

7. Personification of a brain

We, living in the twenty-first century, will know that the above argument is not a mere abstract one, but is directly linked to our daily life. Today, for example, it is said that the compulsive behaviour of a patient with obsessive-compulsive disorder (OCD) is carried out not by the patient but *by the person’s brain*. We can foresee an era in which such understanding is generalised to all kinds of human behaviour. Through *the personification of the brain* and the identification of it as what makes us do the things we do, i.e. as the subject of operation, all human beings, being under the control of their brains, will be regarded as ‘unfree in reality’. There is conceptual confusion here – such that a brain and a human are conceived to be different subjects – and the divergence problem remains set aside, but the personification will certainly continue to progress.

What is important is that the progress in our understanding of the brain-human relation will put aside the explication of what (positive) freedom is. A brain can be on the side of freedom, the determining side, only by being personified as a subject that makes us humans unfree, while we leave unclear what ‘freedom’ strictly means. That is, humans become beings that are ‘unfree in reality’ by metaphorically assigning and transferring their freedom to brains, treating freedom as something that has

been possessed by humans and yet has been unexplained.

This transferring rests on the conceptual ambiguity of or the deferral of accounting for the free-unfree relation *qua* the determining-determined relation. Freedom transferred here corresponds to the positive freedom of choosing one from different possibilities. This is why humans will be regarded as being on the determined side, being made to choose between different possibilities. However, freedom in that sense cannot be temporally located. The essential point about the personification of the brain is this: the personification prevents the divergence problem from arising, by portraying the brain as something that has the feature of invisibility, just as another person has this feature for the second-person freedom.

It is obvious that the structure transferred here can only be given meaning by humans. And, humans give meaning to the structure by resting on the conceptual ambiguity of it. Something (i.e. positive freedom) that has been possessed by humans and yet has been unexplained will be metaphorically transferred to natural objects other than human beings, e.g. to brains, and it will be revealed that what has been transferred is an obscure thing that cannot be temporally located. Then, the viewpoint from the determining-determined relation will vanish. For when the conceptual source of positive freedom dries up, we will no longer be able to transfer our positive freedom to brains, and to understand what is meant by our brain *making us do* something.

Let us clarify the meaning of what has just been said by consulting Libet's 'veto' experiment (e.g. Libet, 1985; Libet, 2004). From this experimental result, Libet draws the following inference: we might not be able to consciously start an action (for the conclusion of Libet, 1985 was that there is the particular pattern of brain activity 'RP' for an action 400 msec before conscious will), but there is room for us to 'veto' an action during the 150 msec between the time of conscious will and the time of performing the action, and so there is also room for our freedom.

So far as we (the authors) know, the evaluations of this 'veto' experiment are generally critical. The design of the experiment is obscure especially with respect to the instruction for 'vetoing'. Moreover, it is natural from the physicalist point of view to suppose that there is also subliminal brain activity preceding the process of 'vetoing'. However, we will not repeat such criticism here. Instead, we will focus on the below point.

Libet, in his inference, aims to retrieve the right of choice from brains and return it to humans; even if a brain *does* attempt to make a human being perform a certain action, a human being can veto the action immediately before the performance of it, thus becoming a being that *does* choose. However, as we saw in the above discussion, it is untenable to adopt the framework in which the *active cause* located at a particular point of time chooses one of the future possibilities. Nevertheless, Libet's 'veto' experiment in fact raises the question of whether it is RP or the veto that counts as

the active cause. This obscures the ‘veto’ experience even more.

When Libet says, ‘The awareness of the decision to veto could require preceding unconscious processes, but the *content* of that awareness (the actual decision to veto) is a separate feature that need not have the same requirement’ (Libet, 2004, p. 147, our italics), he is in the midst of the divergence problem. The ‘content’ is described as something that occurs with no dependence on preceding conditions, and as something due to the occurrence of which future possibilities are chosen. However, as we saw, anything which has such a ‘separate feature’ can only be a one-off chance occurrence. So long as what Libet calls the ‘content’ is something other than a chance occurrence – as he would certainly say – it cannot be temporally located on the divergent diagram bifurcating into vetoing and not vetoing.

People today show various sorts of ethical hesitation in accepting the idea of improving their society by manipulating their brains, because they still do not believe that they are ‘unfree in reality’. That is, they believe that humans have freedom to decide whether to perform an action of medically operating on a brain or of ethically permitting a brain operation (*ignoring* the fact that such an action would also result from the activities of a surgeon’s or politician’s brain). The view that a better society can be realised by medically acting on the criminals’ or neurotic patients’ brains has its value only if there is a subject who can *freely* manipulate brains, and, especially, only insofar as a human being can be regarded as such a subject.

We can now say as follows using the concept ‘*afree*’ (cf. Section 4). If *all* humans come to regard themselves as ‘unfree in reality’, the viewpoint from the determining-determined relation will vanish, as we have said. Then, there will be no subject of free manipulation, that is, all humans, all brains, and all other things will *exist as afree* – rather than unfree – so that only single-line determination or one-off chance can be the description of the reality of the world. Personification of the brain, in essence, arises only in a transition period from an unfree world to an *afree* world. And, although this is an odd expression, personification of a *person* who can be either of the relata of the determining-determined relation also arises only in this transition period.

As science and philosophy progress, they will begin by placing humans on the unfree side by discovering the factors that *seem to* manipulate humans (such as Libet’s RP). But if such studies progress further, they will, through repeatedly confronting the divergence problem (e.g. in the way Libet’s ‘veto’ experiment faces difficulty), slowly bring humankind closer to the *afree* side each day. This is also a process through which humans will gradually become closer to beings that are not humans – that is, for better or for worse, to creatures that are quite different from humans as we know them now.

However, presumably our species will never fully move from an unfree world to an *afree* world. For we humans live in the world of second-person freedom, and the

key to an element of the practice of second-person freedom is to confabulate the active cause as existing in the invisible realm of another person, and to believe that the active cause is temporally located somewhere on the imaginary modal divergent diagram. This confabulation can be an intersubjective illusion arising from ‘cognitive error’. But if this claim is understood as the claim that humans are in fact unfree beings manipulated by something, the discussion of this paper has not been understood sufficiently. To awake from the intersubjective illusion means to go outside the determining-determined relation, so that the very meaning of such expressions as ‘unfree beings manipulated by something’ vanishes.

Acknowledgements

We thank the referees for their useful remarks. This work was supported in part by MEXT KAKENHI Grant Number 26119520.

References

- Aoyama, T. (2010). Are actions and events orthogonal? (Koui to dekgito wa chokkou suruka, in Japanese). *Annual Review of the Philosophical Association of Western Japan (Nishi-Nihon Tetsugaku Nempo, in Japanese)* 18, 87–101.
- Dennett, D. (2003). *Freedom Evolves*. Viking Penguin.
- Horwich, P. (1987). *Asymmetries in Time: Problems in the Philosophy of Science*. MIT Press.
- Johansson, P., Hall, L., Sikström, S. and Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science* 310, 116–119.
- Kane, R. (1999). Responsibility, luck, and chance: Reflections on free will and indeterminism. *The Journal of Philosophy* 96-5, 217–240.
- Lewis, D. (1981/2003). “Are We Free to Break the Laws?,” in *Free Will* (Second Edition), ed. G. Watson (Oxford Readings in Philosophy), 122–129.
- Lewis, D. (1986). *On the Plurality of Worlds*. Blackwell Publishers.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioural and Brain Sciences* 8, 529–566.
- Libet, B. (2004). *Mind Time: The Temporal Factor in Consciousness*. Harvard University Press.
- Loux, M. J. (1979). “Introduction: Modality and Metaphysics,” in *The Possible and the Actual*, ed. M. J. Loux (Cornell University Press), 15–64.
- McTaggart, J. M. E. (1908). The unreality of time. *Mind* 17, 457–474.
- Mino, T. (2008). “Determinism and Freedom,” (“Ketteiron to Jiyuu”, in Japanese) in *Iwanami Lecture Philosophy 2 (Iwanami Kouza Tetsugaku 2, in Japanese)*, eds. T. Iida et al. (Iwanami Shoten), 161–186.
- Nietzsche, F. (1886/1968). *Jenseits von Gut und Böse, Nietzsche Werke: Kritische Gesamtausgabe* VI-2, eds. G. Colli and M. Montinari (Walter de Gruyter), 3–255.
- Nisbett, R. E. and Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84, 231–259.

- Price, H. (1996). *Time's Arrow & Archimedes' Point: New Directions for the Physics of Time*. Oxford University Press.
- Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy* 48, 187–211.
- Terkel, S. (2001). *Will the Circle Be Unbroken*. Thorndike.
- van Inwagen, P. (1975). The incompatibility of free will and determinism. *Philosophical Studies* 27, 185–199.
- Wittgenstein, L. (1953/2001, The German Text with a Revised English Translation). *Philosophische Untersuchungen (Philosophical Investigations, in English)*. Basil Blackwell.

(Received 2013.7.29; Revised 2014.7.31; Accepted 2014.8.11)