JOURNAL OF APPLIED ETHICS AND PHILOSOPHY

Center for Applied Ethics and Philosophy
Hokkaido University

vol.14

February 2023

ISSN 1883-0129

Journal of Applied Ethics and Philosophy

Printed in Japan

ISSN 1883 0129 (Print)
ISSN 1884 0590 (Online)

All queries should be directed to:
The Editor-in-Chief, Center for Applied Ethics and Philosophy
Faculty of Humanities and Human Sciences
Hokkaido University
N10 W7, Kita-ku, Sapporo 060-0810, Japan
jaep@let.hokudai.ac.jp

CONTENTS

Editorial Note

The *Journal of Applied Ethics and Philosophy* is an interdisciplinary journal that covers a wide range of areas in applied ethics and philosophy. It is the official journal of the Center for Applied Ethics and Philosophy (CAEP) at Hokkaido University. The aim of the *Journal of Applied Ethics and Philosophy* is to contribute to a better understanding of ethical and philosophical issues by promoting research in various areas of applied ethics and philosophy, and by providing researchers, scholars and students with a forum for dialogue and discussion on ethical and philosophical issues raised in contemporary society. The journal welcomes original and unpublished regular articles and discussion papers on issues in applied ethics and philosophy.

Nobuo Kurata
Editor-in-Chief

Many thanks to *Journal of Applied Ethics and Philosophy* reviewers

The *Journal of Applied Ethics and Philosophy* would like to thank the following individuals for generously reviewing manuscripts for us between February 2022 and January 2023. The support and expertise of these professionals promote and maintain the high quality of the journal's content. Thank you very much.

Yutaka Kato
Tomohiko Kondo
Katsunori Miyahara
Takuya Niikawa
Katsuhiko Sano
Tsutomu Sawai
Tomoyuki Yamada

Some Perspectives on Moral Status of Human Brain Organoids

With Focus on Consciousness and the "We" as Humans

Yoshiyuki Yokoro
JSPS Postdoctoral Fellow
Institute for Social Ethics, Nanzan University

Abstract and Keywords

Human brain organoids are three-dimensional structures that reproduce the structure and function of the human brain *in vitro*. In this paper, I focus on the question of the moral status future human brain organoids will occupy. It is generally believed that the moral status a being occupies depend on sentience it possesses. However, in this paper, I argue that the presence or absence of a "consciousness" other than sentience may also have moral importance in the light of the features that seem to be unique to human brain organoids, and assert that a new perspective on whether human brain organoids belong to "we" or not may be necessary for thinking abstractly about how they should be treated morally. I argue that even if future human brain organoids lack sentience, they can occupy a moral status when they have so-called phenomenal or access consciousness. On the other hand, I argue that as long as future human brain organoids are as much a part of the "we" in terms of family membership as patients in a persistent vegetative state are. This would imply that even if human brain organoids do not have any kinds of consciousnesses, they may occupy some moral status.

Keywords: organoids, moral status, consciousness, family membership, speciesism

Introduction

Human brain organoids are three-dimensional structures that reproduce the structure and function of the human brain *in vitro*. Recent advances in science and technology have raised concerns about how to treat them morally. Are human brain organoids morally important in their own right? If so, do we have a moral obligation to treat them fairly? How should we think about the rights and wrongs of harming or hurting them relative to humans and non-human animals? The question of the moral status of human brain organoids, somewhat sensationally described as "miniature brains," looms over us as a practical issue that cannot be ignored.

It seems safe to say that the prevailing view amongst contemporary analytic ethicists of a certain stripe, is one according to which the moral status a being occupies and, therefore, the restrictions imposed on its moral treatment depend on the psychological capacities it possesses, especially sentience, which is related to the ability to suffer or feel pain. Following this view, the presence or absence of sentience is essential when considering the moral status of human brain organoids.

However, in this paper, I examine the moral status of human brain organoids by introducing two perspectives that cannot be captured by the prevailing view. The first perspective focuses on the psychological capacity of consciousness of various kinds, and the second on the "we" humans as a moral group. With respect to the first perspective, I argue that the presence or absence of a "consciousness" other than sentience may also have moral importance in light of the features that seem unique to human brain organoids. Even if future human brain organoids lack sentience, the kinds of psychological capacities they will develop matter morally since they can occupy a moral status even when they have only so-called phenomenal or access consciousness. However, in terms of the second perspective, I assert that, the moral status of human brain organoids depends on their being members of the human species and,

thus, members of the "we" as a moral group. As long as we cannot exclude the view that future human brain organoids are as much members of the "we" in terms of family membership as patients in a persistent vegetative state are, and that "we" are nothing less than human animals, even if human brain organoids do not have any kinds of consciousnesses, they may occupy some moral status in virtue of their being members of the "we." This possibility may provide an opportunity to develop a new perspective on whether human brain organoids belong to the "we", in addition to the previously pointed out perspective of whether they have "consciousness" or not, when thinking abstractly about how they should be treated morally.

The remainder of this paper is organized as follows. In Section 1, I briefly summarize the current state of organoid research and identify moral concerns that may arise in the future. Section 2 summarizes the idea of moral status in general, focusing on sentience. Section 3 presents one perspective on the moral status of human brain organoids, considering the variety of consciousnesses. In Section 4, I propose another perspective, based on an analogy involving patients in a persistent vegetative state, that the moral status of human brain organoids can be understood from a new aspect of "we" in terms of family membership. In Section 5, I will briefly reply to two objections to my proposal to include human brain organoids as a member of the "we."

1. Current Status and Future Concerns of Organoid Research

So-called organoids refer to various types of pluripotent stem cells cultivated in dishes or *in vitro* and are specifically three-dimensional biological structures that self-organize to mimic actual organs. In 2008, a research group in Japan's RIKEN Institute showed that cortical tissues derived from pluripotent stem cells self-organized into three-dimensional structures and the characteristics of human cortical development could be reproduced (Eiraku *et al*. 2008). At this point, they were very small in size and had a limited range of neurons, but in 2013, by improving protocols and enhancing nutrient absorption, a group of researchers succeeded in generating three-dimensional structures with cerebral cortical regions displaying an organization similar to the developing human brain at early stages (Lancaster *et al*. 2013). This generation of brain organoids with larger sizes and a wider range of neuron types was a breakthrough, and many human brain organoids (HBOs) have been generated in the laboratory since then.

As evidenced by a multitude of biotechnology examples, there have been remarkable technological advances and achievements. A recent study found that cerebral organoids cultivated for long periods of time acquire the structural features of mature neurons, including spontaneously active neuronal networks and differentiation of photoreceptor cells with light-responsive proteins (Quadrato *et al*. 2017). In addition, certain long-term cultured cerebral organoids derived from human embryonic stem cells have been found to contain specific mature inhibitory and excitatory neurons (Matsui *et al*. 2018). Other studies have demonstrated that cortical organoids spontaneously develop periodic and regular electrical activity, similar to the electroencephalogram patterns of preterm infants, as well as complex neural activity (Trujillo *et al*. 2019).

This series of organoid studies is now considered useful for modeling diseases, testing drug responses, personalized medicine, and organ transplantation. For example, the HBO has been used to identify how the zika virus causes microcephaly in fetuses and to screen for drugs that might weaken the infection (Qian *et al*. 2016). In the future, organoids could greatly advance the development of treatments for neurodegenerative diseases such as Alzheimer's and Parkinson's; organoids may also replace damaged organs, including parts of the brain, without rejection. Progress in organoid research will bring a ray of hope to a number of patients.

However, organoid research is just beginning to develop. In fact, the HBO is still currently much smaller than a real human brain and does not even have blood vessels. It is far inferior to the actual human brain in terms of the number, complexity, and maturity of its neurons, and does not even have the developed sensory inputs and motor outputs necessary to interact with and respond to the external environment. Therefore, it seems unlikely that the current HBO has thoughts or a sense of pain; hence, at the moment, it is an entity that does not resemble us. It is true that we should not deny the possibility that some of current HBOs have some phenomenal conscious experiences, depending on how consciousness is viewed.[1] This means that moral concerns might be raised in the creation and handling of HBOs even at the present time as long as current HBOs as well as future HBOs are regarded as having some consciousness.

However, a more obvious and pressing moral concern

1 For example, if some current HBOs not only exhibit complex neural activities but are also able to have causal interaction with the environment, and thus acquire integrated information, then they have consciousness according to integrated information theory. Since there are many competing theories that offer an explanation for the nature and source of consciousness, it cannot be said that HBOs are free of any ethical concerns even at this time. For a more detailed analysis of HBOs and consciousness, see Niikawa *et al*. 2022.

would arise if HBOs are found to be conscious in most promising theories of consciousness. Given the rate of progress in the field, it is quite likely that the current limitations will be overcome in the near future, and it is not inconceivable that HBOs that interact with their surroundings and eventually gain consciousness will become a reality regardless of which promising theory is chosen. As some ethicists have noted, recent developments "do suggest a growing need to think about the possibility that brain organoids might one day acquire consciousness" (Koplin and Savulescu 2019, 760). Indeed, the International Society for Stem Cell Research 2021 cautioned that the ethical issues could arise, as organoid research has progressed (International Society for Stem Cell Research 2021). Thus, we can predict that concerns about how to deal morally with HBOs and other organoids will sooner or later become one of the most pressing issues that ethicists must address.

2. Moral Status and Sentience

What moral status will future organoids, including HBOs, occupy? To address this question, the concept of moral status must be clarified. In general, when a being has characteristics related to moral values and occupies a moral status, some moral restrictions are imposed on its treatment. Among other things, moral agents have an obligation to treat those who occupy a moral status fairly, that is, without harming them, as there are strong moral reasons against harming or killing beings with moral status (Warren 1997, 3–4). In other words, occupying a moral status means that one has certain rights and, thus, matters morally.

As some theorists have argued, for a being *x* to occupy a moral status means that *x* is morally important in its own right, without depending on other beings (Harman 2007; Kamm 2007, 227–28). In short, moral status is closely tied to the intrinsic moral value of the being that occupies it. A representative example of a being with such an intrinsic value is a person. For example, consider the case of the person called Claire. Claire occupies a sufficiently high moral status, which is a reason for someone to refrain from harming her, except in special circumstances such as self-defense. On the other hand, the reason for refraining from harming Claire's car is not because of the moral status that the car itself occupies, but rather because harm to the car indirectly harms Claire, and Claire herself occupies a moral status. In other words, we should refrain from harming her car only because the car has a special relationship with her (in this case, a relationship of ownership), and not because the car itself is a being that morally requires special treatment. In this sense, objects such as cars do not occupy any moral status and are not morally important in themselves.

How, then, can we determine whether a being occupies a moral status? There has been no small amount of advocacy for the idea that a being's "consciousness," especially sentience, which is a sensation regarding the qualitative experience of pleasure and unpleasure, such as pain and suffering, contributes to the conferral of moral status (Singer 1979; DeGrazia 2008). According to one of its proponents, David DeGrazia, sentience is phenomenal consciousness plus hedonic valence, which can make things better or worse for the subject, depending on the perceived quality of the experience.[2] For example, let us say that I hit a dog with a hard stick. It is clear that my action is morally despicable, even more so if the dog is my neighbor's dog; however, even if it is a stray dog with no special relationship to anyone, my action still deserves strong condemnation because dogs can feel pain, and the pain, which is harmful to the dog, deserves moral consideration. Therefore, I have a moral obligation not to inflict pain on dogs. In general, any creature that can feel pain or suffer is, by the very fact, morally significant and occupies some moral status. On the other hand, if I had hit a non-sentient stone on the road instead of a dog, there would have been no moral problem because the stone cannot experience pain or suffering. Accordingly, sentience, which relates to the ability to suffer and experience pain, is a major criterion in determining whether one has a moral status.

Let us now return to the moral status of future organoids (and, for the sake of brevity in the discussion below, "organoids" will be taken to refer only to cultivated human stem cells). If the widely accepted explanation above is correct, then among organoids, only HBOs raise moral concerns. For example, liver and kidney organoids are unlikely to have morally important features such as sentience, and therefore do not occupy a moral status. Therefore, a series of organoids that mimic human organs will probably not raise serious ethical concerns.[3] However, the situation is different for HBOs. The closer the future HBOs are to the actual human brain in terms

2 Underlying these ideas is the assumption that having the capacity for sentience leads to having interests (because it leads to experiential well-being) and that those who have interests occupy a moral status. Alternatively, suffering as an experience of pain with a dislike of such a sensation can be considered intrinsically bad in itself because of its phenomenal character. See Kahane 2009, on this.

3 This does not mean that no regulation is needed for the handling of a series of organoids other than brain organoids. For example, some legal regulations may be necessary to prevent their commercialization of them for transplantation and the resulting abnormal price increases. It should be noted that interview surveys with non-experts have shown that HBOs are of a greater moral concern than other organ organoids. See Bollinger *et al*. 2021, on this.

of function and structure, the more likely they are to have sentience. There is no good reason to exclude this possibility. If this is the case, then sentient future HBOs, like dogs, should occupy some moral status, and some moral restrictions should be imposed on their handling. For example, it is morally imperative to refrain from inflicting unnecessary pain on sentient HBOs in the course of experimentation, as well as to avoid choosing disposal methods that cause them to experience excessive suffering.

However, two points must be noted. First, it is expected that there will be many cases where it is difficult to determine whether HBOs have sentience. In such cases, it seems morally justified to follow the so-called precautionary principle and treat them as sentient beings, taking a generous stance for the time being. This is because the harm of mistakenly treating a sentient being as a non-sentient being is overwhelmingly greater than the harm of mistakenly treating a non-sentient being as a sentient being. This is similar to how it is worse to wrongly convict an innocent person of guilty in criminal justice than to wrongly convict a guilty person of not guilty; hence, the principle of the presumption of innocence is upheld (Sebo 2018).

Second, future HBOs' sentience and thus occupation of some moral status do not immediately prohibit scientists from using or disposing of HBOs in their experiments. From a consequentialist point of view, if HBO research and experimentation have significant benefits, such as the promotion of disease modeling and personalized medicine, then those benefits may outweigh the moral status of sentient HBOs in some cases. Nevertheless, as with animals in animal experiments, we cannot unconditionally do anything to HBOs. As long as HBOs occupy a certain moral status because of their sentience, the amount of pain and suffering they endure should be minimized, or at least not neglected, unless there is some particular reason.

3. Varieties of Consciousness

The discussion thus far has revealed that future HBOs will occupy a certain moral status as long as they have sentience. However, this is merely an application of the moral status argument mainly constructed for non-human animals to HBOs. The above argument does not reflect the unique characteristics of HBOs. Then, what exactly are the "unique characteristics" of HBOs?

In my opinion, one feature is that HBOs may or may not have various kinds of "consciousness" in some cases. In general, there is no doubt that the everyday word "consciousness" is polysemic. Ethicists often regard the word as interchangeable with sentience, but both so-called phenomenal and access consciousnesses are also types of "consciousness" (Block 1995). Here, it is beneficial to briefly review the terminology.[4] Phenomenal consciousness is a subjective experience of qualitative content that concerns what it is like to see or feel something. This subjective experience is essentially a first-person, private phenomenon that no one but the being who actually has it can experience. For example, my experience of tasting a lemon is not identical to yours, even if they are qualitatively similar. Access consciousness, on the other hand, is the cognitive, intentional, and functional aspect of consciousness associated with the availability of information. When used as a premise for reasoning or for rational control of action, the subject with that mental state can be said to be in a state of access consciousness, characterized by representational content. Beliefs, desires, plans, and intentions are typically classified under this type of consciousness.

The problem is that if at least two of these "consciousnesses" are characteristics that are related to moral importance as well as sentience, then the moral status of HBOs in the future will take on a multifaceted aspect. Among other things, when HBOs have *this* consciousness but not *that* consciousness, how they should be dealt with it morally becomes much more complicated.

Consider, for example, a being that lacks sentience but has phenomenal consciousness, that is, a being that must have some subjective experience from a first-person perspective but has no sensation of pleasure or unpleasure about it. It is possible that future HBOs that develop only specific nervous systems or brain regions will be examples of such beings. Is it then plausible that the HBO can subjectively see the color of and smell some chocolate in front of it but has no moral status because it has no sentience? This argument is debatable. There is still room to think that having a set of phenomenal consciousnesses makes one's life worth living, and that depriving one of such first-person consciousnesses in any way is morally unacceptable. For example, assume that the dignity involved in a non-discriminating kind of respect lies in a certain norm associated with the phenomenal privateness that is characteristic of phenomenal consciousness, namely, the norm of incorrigibility, according to which no conscious subject can be corrected or interfered with by another subject about what goes on in their mind. According to Uriah Kriegel, it is because each person's unique

4 There are, of course, many other kinds of "consciousness" that cannot be dealt with in this paper due to space limitations. Specific examples of these are self-consciousness, monitoring consciousness, reflective consciousness, higher-order consciousness, minimal consciousness, and so on.

subjective experience entails the norm of incorrigibility that a phenomenally conscious being becomes worthy of being respected—by virtue of its inherent dignity and being treated as an end (Kriegel 2017). At this point, a phenomenally conscious life is better than a phenomenally non-conscious life by itself and might be morally important thanks to having the dignity. There are many kinds of phenomenal consciousnesses, including those related to perception, the sense of smell, and emotions; however, whatever the kind, maintaining phenomenal consciousness enhances the value of the owner and losing it by being killed or maimed would be morally undesirable as it would be harmful to the owner who are worthy of being respected.

Of course, it is not without question whether phenomenal consciousness itself is morally importance. This is because when comparing a world where there are only minimally conscious organisms that have the sole experience of brightness with a world where there are only organisms that have no conscious experience at all, it is not at all clear whether the former world is really better (Lee 2019). In addition, cases of the separation of phenomenal consciousness and sentience might be unlikely in organisms because of natural selection (DeGrazia 2021). In the case of HBOs, however, it is not impossible to adjust the brain to reproduce only certain brain functions relevant to phenomenal consciousness; therefore, the existence of HBOs that can subjectively experience things without sentience is one of the more realistic possibilities.[5] Moreover, as the subjective sensory experiences that each HBO can have as an only owner of them become richer and broader, it is at least undeniable that each HBO, as a phenomenally conscious entity, becomes a being worthy of respect and thus gains the status of morally important being in its own right. Therefore, insofar as phenomenal consciousness itself can be considered to have some moral importance, or insofar as having rich and varied types of phenomenal consciousnesses is considered more important than a complete lack of phenomenal consciousness, then it follows that HBOs lacking sentience does not necessarily lack moral status.

Various other patterns can be envisioned depending on the polysemy of "consciousness." For example, if future HBOs do not have sentience (and hence phenomenal consciousness) but possess access consciousness, what will their moral status be? According to some ethicists, sentience and phenomenal consciousness do not contribute to moral status on their own (Levy and Savulescu 2009; Levy 2014). What is directly related to moral status is the relatively advanced cognitive capacity to have future-oriented desires about what one's life will be like and to think and believe that one is temporally viable. However, to make full use of these psychological capacities, information must be sufficiently available for rational thought and contemplation, which would require sophisticated access consciousness. Indeed, the states in which we care about and desire our own future contribute to our interest in survival, but they are mostly informational states and not phenomenal states.

Furthermore, a moral emphasis on access consciousness instead of phenomenal consciousness would be compatible with so-called *desire-satisfaction theory*, one of the most promising theories of well-being (Kahane and Savulescu 2009). Under this theory, which holds that the satisfaction of one's desires makes one's life better, having access consciousness involving desires and their satisfaction—even if one does not feel such a satisfaction phenomenally—is a necessary condition for a better life. This means that it is bad for the subject that his or her relatively sophisticated future-oriented desires, such as desires regarding how his or her own life goes, fail to be satisfied by being killed or maimed. Accordingly, having this sophisticated kind of access consciousness, including desires, can be a sufficient condition for the subject to occupy a moral status, as long as there are moral reasons to refrain from interfering with or making impossible the satisfaction of his or her future-oriented desires. If this is the case, then an HBO with a sophisticated level of desires, preferences, beliefs, and other components of access consciousness, even if it lacks any phenomenal experience or sentience, should be said to occupy a moral status, since it could be harmed by being killed because of the unsatisfaction of the future-oriented desires it would have been satisfied if it was alive.

As discussed above, moral concerns arise in conjunction with future HBOs' unique characteristics, not often found in the natural world where natural selection works. Determining the moral status of HBOs is difficult without clarifying what kind of consciousness they can have and how morally important possessing each type of consciousness are.[6] Conversely, considering the moral status of future HBOs necessarily requires a perspective on what kind of consciousness future HBOs

5 Similarly, it has been suggested that it is possible to develop a fully mature brain *in vitro* that is capable of reason but feels no pain at all because it has no nociceptors. See Hostiuc *et al*. 2019.

6 Among other things, since the acquisition of self-consciousness has been considered to be deeply related to the moral status of the fetus in the context of the moral pros and cons of abortion, the moral status of self-conscious HBOs would also be an issue if viewed analogously to this. For more on this, see Koplin and Savulescu 2019; Hyun e*t al*. 2020.

will achieve as a specific issue.[7]

4. Family Membership in Relation to "We"

However, does the perspective on the moral status of future HBOs really require only sentience or various forms of consciousness? Previous studies examining the moral status of HBOs have focused exclusively on the possibility that they possess psychological capacities such as consciousness, reason, and self-consciousness (Munsie *et al* 2017; Hostiuc *et al*. 2019; Koplin and Savulescu 2019; Hyun *et al*. 2020; Lavazza and Pizzetti 2020). I think we need to look at another vector as well, because while ideas that strongly link moral status to psychological capacities do have advantages, such as encouraging moral consideration of sentient or conscious beings, they also run the risk of mercilessly cutting off unconscious beings from the discussion of moral status. Prime examples of the unconscious beings are patients in a persistent vegetative state (PVSs) who are in a so-called coma due to complications of an underlying disease or head injury. Assuming that they are beings with no phenomenal consciousness or sentience at all, let alone access consciousness,[8] it follows that they have no psychological capacities of any moral importance. Nevertheless, our common sense or conscience strongly resists the claim that they do not occupy any moral status and are morally unimportant in their own right. Does this mean that the attempt to explain moral status entirely in terms of psychological capacities, such as consciousness, is itself unreasonable? If that is the case, we will need to introduce another perspective other than psychological capacities when considering how to deal morally with future HBOs.[9]

As Jeff McMahan pointed out, the treatment of PVSs is analogous to the treatment of pets (McMahan 2002, 228–32).[10] There are certainly reasons PVSs should not be treated badly, McMahan argued, not because of their intrinsic moral value, but because of the special relationships (of love and sympathy) they have with their families and relatives that should be respected. He asserted that while our sense of kinship with human PVSs pushes us to treat them well, it also leads us to dismiss many animals as strangers, greatly reducing our sensitivity to them. Giving moral status to PVSs who lack morally important traits—traits that McMahan himself linked to a psychological unity—simply because they are "human" is comparable to pernicious nationalism, which regards non-nationals as inferior and sometimes "dehumanizes" them (McMahan 2002, 221–22).

However, as Eva Feder Kittay argued, drawing a line between those who have moral status and those who do not based on psychological capacities is itself an aspect close to pernicious nationalism (Kittay 2005). The dangers of drawing a line between persons with sophisticated psychological capacities and PVSs, viewing the latter as inferior in moral importance, even for the laudable purpose of promoting animal well-being, cannot be overemphasized. Kittay denied that the presence or absence of certain psychological capacities is the basis for conferring moral status and instead argued that family membership is morally important. In her view, family membership is not pernicious unless it is abused and falls into nepotism; rather, it plays an important role when we need to rely on someone, such as in childhood or when we are suffering from an illness. She expanded the moral group from the family to the human species, to which we all belong, and argued that "membership in a group of moral peers based solely on species membership has as its appropriate moral analogue family membership." In other words, "as humans we are indeed a family" (Kittay 2005, 124). As long as beings such as PVSs and severely cognitively impaired patients are human, even if they lack the psychological capacities in question, they are members of the "we" in terms of a moral group called a family, which gives them proper moral status.[11]

7 Therefore, as some theorists have pointed out, it will be important to introduce tools to evaluate whether organoids have consciousness or not. For example, integrated information theory can be a useful evaluation tool because it can quantify consciousness according to the degree of information integration. See Lavazza and Pizzetti 2020.

8 This assumption is necessary because it is generally not easy to distinguish between PVSs and patients in a state of minimally conscious state (MCSs). Unlike PVSs, MCSs have intermittent consciousness and may also experience pain.

9 This non-psychological perspective is often overlooked when considering the ethics of human brain research. For example, Henry Greely collectively calls things such as HBOs and genetically edited non-human animals "human brain surrogates" and examines their welfare and rights (Greely 2021). In doing so, Greely focuses on the pain and consciousness that these surrogates can have, while asserting that they are "not human beings" (ibid., 38). However, the basis for the assertion is not clear, and furthermore, Greely unjustifiably excludes the option that membership of the "we" as human beings might itself be of moral importance, and thus lead to a better understanding of the welfare and rights of these surrogates.

10 In fact, McMahan's focus here is primarily on severely retarded human beings, but I will use PVSs for our discussion because there are few notable differences between severely retarded human beings and PVSs in terms of moral treatment.

11 On the other hand, Kittay cites the capacity for caring, empathy, and kindness, as well as the capacity for rationality

But here, a question might arise. Why is being a member of the "we" morally important enough to confer moral status on an entity? Kittay does not say anything on this point, but in my view, the moral importance of membership of the "we," which is morally analogous to family relationships, stems from an understanding of the nature of morality. We usually place moral importance on "fellowship," such as family or ethnicity. These particular social or biological relationships provide a moral reason for preferring one being over another. If we consider this on a global scale, being a human being is much like being a member of a family. In other words, we prioritize humans over nonhuman animals because of a sense of belonging to the human species as a sense of solidarity with our fellow human beings. This view is based on a focus on the so-called *agent-centeredness of morality*. To take a simple example, if two people, one a child and the other a stranger, are drowning and only one of them can be rescued, it is morally permissible for the mother to give priority to rescuing her own child. This conclusion is because morality does not necessarily require an impartial observer or agent-neutral moral point of view, but can be considered agent-centered or agent-relative because moral goodness and wrongness are located in the attitudes of agents, not in actions. In the above case, the mother is considered to have a special moral obligation to her child regardless of any moral reasons the other agents might have.

If so, it is not so far-fetched to extend this to humans as a whole and consider that we have a special moral obligation to those who belong to our own species. A similar point has been made by Bernard Williams, who like Kittay, argued that humans could be morally more important than other animals. To borrow his words, "there is no other point of view except ours in which our activities can have or lack a significance" (Williams 2006, 137) and thus "creatures are treated in one way rather than another simply because they belong to a certain category, the human species" (ibid., 142). Under this view, taking morality per se from nowhere or from a cosmic perspective is itself a source of error. Human beings are morally important from our perspective as human beings. Therefore, the membership of the "we" confers a moral status because moral importance is linked to a sense of solidarity or fellowship with the human species as seen in the family on a global scale.

Let us now return to the discussion of the moral status of HBOs. I think that Kittay's suggestion hits the nail on the head and has serious implications for the moral status of HBOs. Now, if we understand the "we" in terms of a moral group called a family, then for HBOs to be one of us, insofar as they fall into the same category of the human species under which we all fall—even if it lacks any morally significant psychological capacities—that classification would give us reason to believe that HBOs occupy at least the same moral status as PVSs. This is because, via the relationship of family membership, a member of the human species is associated with some moral status, and insofar as HBOs become members of the human species, they become members of the "we" and have a certain concomitant moral importance.[12]

However, can the future HBO really be a member of the "we" in terms of a family membership and thus a member of the human species, even though it is only a brain? To examine this claim, a view related to the metaphysics of "we" may be helpful. The view, called *animalism*, emerged at the end of the 20th century and has since gained a certain amount of support (van Inwagen 1990, 142–81; Olson 1997, 124–53; Merricks 2001, 118–37). According to this view, we are all human animals as living organisms called *Homo sapiens*, and you and I are each numerically identical to the corresponding specific human animals. On the contrary, the existence or non-existence of rationality, self-consciousness, and other abilities thought to be unique to the concept of a so-called "person" has nothing to do with our essence and identity; rather, what we are consists in the biological fact that we are human animals and each of us human animals are identical diachronically if and only if we continue to sustain our individual lives as self-organizing biological and chemical events that maintain a complex internal structure. As some animalists have agreed, we will disappear from the world when we cease to be a human animal, that is, when we lose our life-sustaining functions and die a biological death.[13]

Notably, the dependence of our essence and identity

which is characteristic of personhood, as important properties for morality. I do not deny that these abilities may play an important role in acquiring moral status, but in light of the fact that PVSs do not even have these abilities, I will ignore them in this paper.

12 In a similar vein, Thomas Scanlon, from the perspective of contractualism, cites relationships such as friendship and family as deeply related to the requirement of right and wrong, and also emphasizes the analogous relationship of being a member of the human species. See Scanlon 1998, 183–187, on this.

13 Not all animalists agree with this proposal. For example, some animalists claim that we cease to exist completely only when the structures appropriate to life are destroyed (e.g., most of our parts become dust), which in turn allows us to continue to exist not only as life-preserving animals, but also as lifeless or dead animals (Mackie 1999; Snowdon 2014, 114–20). However, the question of what is a human animal as a living organism is a serious philosophical question in itself, and is beyond the scope of this paper. At the very least I am assuming here a standard animalism according to which the nature and identity of the human animal lies in the preservation of its life-supporting functions.

on being human animals does not negate the fact that we humans can be reduced to the size of a brain. Just as the loss of a limb does not necessarily mean the loss of life, that is, our extinction, even if we lose not only our limbs but also our torso and are pared down to just our brain, we can continue to exist as long as the brain continues to serve as a controlling center for the autonomic nervous system (through external life support systems, etc.) and continues to sustain life. As one of the main proponents of this view, Peter van Inwagen pointed out that "the severed head is a genuine living organism" (van Inwagen 1990, 177).[14] In the view of animalists like van Inwagen, there is a crucial difference between the brain and the rest of the body.

It would be beneficial to consider this difference for a moment. On the one hand, the brain only requires a very simple pump to keep the brain supplied with blood to keep it functioning. Even if a certain part of the brain, such as the cerebrum, which controls advanced mental abilities like reasoning and memory, is not functioning, if the brain can perform functions essential for life in a certain environment, it can be said to continue to sustain its life on its own. On the other hand, the rest of the body including the torso and limbs need a constant supply of electrically transmitted information to keep functioning and maintain itself. In other words, the rest of the body cannot coordinate its activities without something that plays a role in sending down very complex electrical information into the spinal column, and the brain as a controlling organ does play exactly such a role. If the dividing line between whether something is an organism or not is considered to be whether it can control something on its own initiative and thus maintain itself, a being with no torso or limbs but only a brain, unlike the rest of the body, can be a human animal as a living organism, i.e., a member of us. Again, in van Inwagen's words, we can summarize it as follows: "We should remember that the brain is not only the seat of our experience and our conscious thought; it is also the seat of our capacity to have lives" (*ibid*., 179).

If these claims of animalism are correct and if being a brain-sized being does not preclude belonging to human animals, then there is no reason to exclude the possibility that future HBOs will acquire biological life and belong to the human animals. When there can be a brain-sized "we," it is understandable that a being born from human stem cells and with a brain size structurally and functionally analogous to that of a human brain could also be a human animal, that is, a member of the "we," even if it is not born from the womb and requires external life support. It is true that we do not know exactly what kind of biological life an HBO can acquire, and it is not easy to capture the exact moment when an HBO acquires life. Nevertheless, it can be said that HBOs that belong to the biological category of humans without so-called bodies may exist in the future for reasons similar to how we humans can be beings pared down to the size of a brain. And here, applying Kittay's abovementioned suggestion, it follows that when a future HBO belongs to the human species and becomes one of us, the individual will be included in the family membership that allows for the granting of a moral status. This leads, analogously to the PVSs case, to the assertion that future HBOs belonging to our own species of human animal can be morally important in their own rights, even if they lack psychological capacities such as sentience and various forms of consciousness.

Thus, I argue that as long as the moral status of PVSs depends on their being one of the "we" in terms of family membership, and we humans can be pared down to the size of our brains, the moral status of future HBOs must be considered in terms of whether they are one of the "we." To the same extent that it would be morally unjust to "dehumanize" PVSs and treat them like objects by depriving them of their moral status, it may be morally impermissible to "dehumanize" some HBOs that lack consciousness or other psychological capacities and treat them like objects. Such a suggestion has not been made by ethicists before, but it seems to me that the perspective based on the analogy with PVSs might illuminate how HBOs should be treated.

5. Objections and Replies

When considering the moral importance of HBOs, it is not enough to consider some psychological capacities; we must also consider whether HBOs belong to the human species and thus to the "we" in terms of family membership. However, there may be some objections to these arguments. Before concluding this paper, I would like to address two possible objections and try to respond to them.[15]

14 See also Olson 1997, 44–46; Merricks 2001, 52 for similar remarks. As Derek Parfit has pointed out (speaking for animalism), when one assumes the existence of an organism whose head or brain occupies 90% of its entire body, it is relatively natural to assume that the identity of the organism depends on the identity of its brain. If this is the case, it is hard to deny the possibility that the brain plays an important role in the identity of animals like us human animals that have brains, regardless of what percentage of their bodies is occupied by the brain. See Parfit 2008.

15 There may be more than two objections to my arguments. In particular, as an anonymous reviewer of this journal suggested, if being a member of the "we" is viewed in terms of human community and interaction with us in an environmental space, HBOs who do not have bodies to

The first objection concerns understanding moral status. Moral status was introduced earlier as related to intrinsic value. According to this, the moral status that *x* occupies means that *x* has moral importance or value on its own, without dependence on other entities. However, family membership, based on being one of the "we" belonging to the human species, is clearly a relational property and is not intrinsic to HBOs. Is it possible for such a non-intrinsic element to support the intrinsic value of HBOs?

I would like to answer this question in the affirmative. The analogy to be drawn is based on other types of value, especially aesthetic value. Suppose an artifact has no intrinsic properties that contribute to its aesthetic value. It is commonplace and unremarkable in terms of form, color, and texture. However, in my opinion, such artifacts can, in some cases, have intrinsic aesthetic value as works of art, because they can become objects of aesthetic criticism through social relationships with viewers. A good example of this is Marcel Duchamp's Fountain or other avant-garde modern art. Although it might not have the intrinsic value of powerfulness that Michelangelo's David has, it is a work of art in its own right, as long as it is placed in an appropriate art-world space, and the viewer perceives it as a work that overturns conventional aesthetic concepts. In other words, whether an artifact has intrinsic value as a work of art is not necessarily determined by its intrinsic nature alone.

Similarly, even if an HBO does not have any morally significant intrinsic properties such as sentience or consciousness, it can still have intrinsic moral value and occupy a moral status through some non-intrinsic relations. What I have argued in the previous section is that one of those relations is equivalent to being one of the "we." In other words, HBOs can have intrinsic moral value in virtue of its family membership along the rest of us humans and hence the relation of being one of the "we" in terms of family membership is a source of the intrinsic moral importance HBOs can have in some sense.[16] However, the next question that must be addressed here is this: What exactly is the "intrinsic value" in question?

Generally, the intrinsic value of *x*, whether aesthetic or moral, is understood as a monadic property that is grounded in (or supervenes on, depends on) the intrinsic property of *x*. Following this, both Duchamp's Fountain and some HBOs as unconscious human animals do not seem to have any intrinsic properties that could be morally valuable by themselves, such as aesthetic powerfulness or psychological properties, and therefore they do not seem to have intrinsic value grounded in those properties either. In my view, however, there is no particular reason why the intrinsic value at issue here should be taken as a monadic property. In other words, there remains the possibility that the intrinsic value that *x* possesses is nothing other than the value-related dyadic relation *R* that obtains between *x* and an entity *y* which can be different from *x*. Since intrinsicality can be established not only for monadic properties but also for dyadic relations, the relation *R* is intrinsic as long as *R* obtains independent of any external entities other than *x* and *y* as its relata. In particular, in the case of moral value, let us recall the rationale on which the moral status of HBOs was thought to be conferred by the membership of the "we." The argument in favor of this was the agent-centeredness of morality, and in particular, the moral importance not from nowhere but from the perspective of us, human beings. In other words, moral status can only be established from the perspective of us, human beings, and people such as PVSs and HBOs are morally important only in relation to us.

Thus, in the case of moral value held by HBOs,

interact with us may not be members of the "we" (See Cordeiro-Rodrigues and Ewuoso 2021 for a relational approach that emphasizes the capacity to positively relate to others). I cannot analyze such views in detail, but it is debatable whether so-called brains in a vat (BIVs) that often appear in philosophical thought experiments, are excluded from the membership of the "we" as a moral group only because they cannot interact with us humans, or do not have bodies to relate to others. I do not currently believe that a BIV, detached from your body and irreversibly losing your consciousness, would lack moral importance if it continued to be yourself because of its life-sustaining function according to animalism. However, I agree with the assertion that a being's interaction with us humans may make the being more morally important. In this regard, there would be no need to deny that bodily PVSs occupy a relatively higher moral status than BIVs.

16 John Robertson described the creation of embryonic stem (ES) cells with the capacity for self-renewal and differentiation into diverse cell types for research purposes as "inherently disrespectful" of human life (except when it would not otherwise be possible to advance important research) because embryos serve as a powerful symbol of human life (Robertson 1999). However, my arguments focus on the moral importance of belonging to us humans, whose essence and identity consist in the biological fact, based on a moral analogy with PVSs and the agent-centeredness of morality, instead of regarding embryos or HBOs as symbols of human life. In other words, I am not committed to the claim that beings with symbolic value should be respected. And, as critics have pointed out, the damage to symbolic value may be relevant to aesthetic value in the sense of "bad taste," but it may not be unethical and thus not lead to damage to moral value (Bortolotti and Harris 2006). The analogy between aesthetic and moral values I use is not intended to emphasize the existence of symbolic values, but to show that there are intrinsic values based on some relations. I would like to thank an anonymous reviewer for bringing these points to my attention.

assuming that it is not, in fact, a monadic property but a dyadic relation *R** that can only be established in relation to us humans, *R** can be "intrinsic" regarding to its relata (i.e., HBOs and us humans) if the obtaining of *R** is grounded in (or supervenes on, depends on) the intrinsic properties of its relata, instead of being dependent on any external entities.[17] Now the intrinsic relation of *R** is the very relation of moral family membership. Given the intrinsic nature of HBOs and us humans, i.e., being human, moral family membership, or being one of us human animals, which is the basis of moral values, obtains between them automatically. This is analogous to the intrinsic relation of being taller; given the intrinsic property of both Claire and Michael, i.e., their height, being taller obtains between them automatically. Just as no other extrinsic elements are required for the obtaining of the relation of being taller between Claire and Michael, so too no other extrinsic elements are required for the obtaining of the relation of moral family membership between HBOs and us humans. Thus, it is not impossible for HBOs to have "intrinsic moral value" (in relation to us humans) as long as the moral value is none other than a dyadic relation rather than a monadic property and the very relation is being one of us in terms of family membership as an intrinsic relation whose obtaining between HBOs and us humans is grounded in their intrinsic property.

Let us now move on to the second objection. This emphasizes the earlier crisis of pernicious nationalism, which has often been linked to "speciesism" (Kuhse and Singer 1990; McMahan 2002, 217–28). The claim that belonging to the human species is a basis for moral importance may not be different from viewing belonging to a particular race as a basis for moral importance. In other words, favoring one being simply because it is a member of the "we" as human animals, while neglecting consideration for other beings, seems to be as abhorrent a prejudice as racism, which decides whether or not to favor a person based on the color of the skin or ethnicity.

The many responses to this objection are impossible to examine fully here (Hopster 2019). However, limiting the scope of the discussion to my proposal, I can provide brief responses from at least two viewpoints.

First, the claim that HBOs are one of the "we" in terms of family membership, enabling the HBO to occupy a moral status, does not imply a lowered moral status for non-members. As we saw earlier, insofar as sentience and various forms of consciousness are included among the features associated with moral value, the moral status of non-human animals should be considered by virtue of their psychological capacities. In some cases, a chimpanzee who has not only sentience but also sophisticated access consciousness may occupy a higher moral status than HBOs or PVSs. There is no reason to deny this at least, and the moral importance of some psychological abilities never diminishes depending on whether these beings are one of the "we" or not.[18] In other words, my proposal is not a call to downplay consideration for non-human species or to prioritize the interests of members of the human group over others; therefore, it is far from any discriminatory actions.

Second, strictly speaking, my proposal is not that the HBO occupies a moral status precisely because it belongs to the human species. It is true that moral status is determined not only by psychological abilities but also by being one of the "we" in terms of family membership. However, it is only through animalism, which is one of the leading metaphysical views, that we can understand that such membership is based on the biological species of human animals. At this point, I strongly disagree with the unwarranted or unconditional equation between the potentially morally significant relationship of family membership and the biological category of belonging to humans. In my view, one of the grounds for moral status lies in family membership construed as being one of the "we," and as a result of my support for animalism, all humans, including HBOs, can occupy a moral status because "we" are nothing other than human animals who can also be brain-sized. This is not to argue that the biological concept of being human animals have intrinsic moral value. It is stating that each of us has intrinsic moral value precisely because we are members of the "we" in a family; hence, future HBOs can also have intrinsic moral value as long as they are human animals like us because "we" can be found to be numerically identical to each human animal in the metaphysical framework of animalism.

On the contrary, options other than animalism may well be possible with regard to what "we" are. For example, it is possible to expand the scope of "we" the family to include all living things that have biological life or limit it to "social humans" in the sense that they are born of human parents, raised by humans, and treated as humans in human society. Alternatively, we could even restrict it to "persons" who possess advanced psychological abilities, such as self-consciousness and rationality. If we limit it to a "person," some HBOs would not occupy any moral status at all, even if they do belong to the class of human animals because beings who do not have psychological abilities are no longer

17 This kind of *intrinsicality* might be called internality. I owe the view that an intrinsic (in my sense) or internal relation is one that depends on the intrinsic properties of its relata to the eminent philosopher David Lewis (Lewis 1986, 62).

18 Peter Singer famously wrote: "pain is pain, and the importance of preventing unnecessary pain and suffering does not diminish because the being that suffers is not a member of our own species" (Singer 2009: 220). I completely agree with this suggestion.

part of the "we." Furthermore, we can even separate "we" as a moral group from "we" as metaphysical beings. In this case, we can consider HBOs as metaphysically belonging to the "we," i.e., human animals, but excluded from the other "we" in the sense of a moral group analogous to the family.

What is important here, however, is not which view is valid but rather that the adoption of some views will reveal moral concerns that have been overlooked. As we saw in Section 3, the moral status of HBOs that do not have sentience but possess phenomenal or access consciousness depends on whether each consciousness really has moral value. Some may hold that sentience is the only morally important feature. However, as long as there can be positions representing other ways of thinking, we cannot claim that there are no moral concerns over HBOs that are not sentient. In the same way, the moral status of HBOs, which lack any morally significant psychological capacities but have developed life-sustaining functions, depends on the claim that other analogous beings, such as PVSs, have a moral status because they belong to "we" the family, and on the claim of animalism, which identifies each of us with each biological human. If these claims are true, and furthermore, if we can think that "we" as a moral group coincides with "we" as beings whose nature and identity should be metaphysically elucidated, then moral concerns would have to be raised for HBOs, which do not have psychological capacities. Put another way, if we avoid the danger of depicting PVSs as lacking intrinsic moral value and expelling them from membership of the "we," and given that animalism is a promising metaphysical view, the claim that unconscious but alive HBOs occupy a moral status should not be ignored as an option.

Of course, it is easy to imagine that there could be many cases of HBO that may or may not belong to the class of humans. An obvious example would be the transplantation of HBOs into a non-human brain, such as a rodent brain. In fact, such transplantation was performed in real life, and in this experiment, a "fusion" occurred between the host tissue and the HBO, and functional neural networks and blood vessels developed in the graft (Mansour *et al*. 2018). It is highly debatable whether this so-called chimeric entity really belongs to the class of humans.[19] However, it seems morally more desirable to treat the HBO in question as a being that belongs to the class of humans and therefore to the "we," unless it can clearly be judged as not belonging. This is just as it is morally legitimate to treat the HBO in question as a sentient being, basically following the precautionary principle, even if it is difficult to determine whether it is sentient, as discussed at the end of Section 2. The harm associated with mistakenly treating a being that belongs to the "we" as one that does not belong is overwhelmingly greater than the harm of mistakenly treating a being that does not belong to the "we" as one that does belong. In other words, it is by no means impossible for an HBO that lacks consciousness to occupy some moral status, even when it is unclear whether it belongs to the class of humans.

Even after the above discussion, I cannot completely deny the possibility that my proposal may result in an abhorrent bias that overemphasizes the importance of belonging to the class of humans. However, I believe that a response from the above two viewpoints can contribute to a more convincing theoretical path for thinking that HBOs can occupy a moral status without having any psychological capabilities.

Concluding Remarks

Let me summarize what I have argued. In my view, the moral status of HBOs in the future demands at least two perspectives related to the characteristics that seem to be unique to HBOs. The first is the varieties of consciousness an HBO can possess. Rarely discussed by applied ethicists, the second is whether the HBO belongs to the "we," i.e., to the class of human animals. Indeed, these two perspectives, even if they point in the right direction, are only rough sketches for an accurate assessment of the moral concerns surrounding HBOs, and they leave many issues and problems untouched. For example, does the distinction between the first and second perspectives imply that there are two types of moral status—or does it only indicate that a unified moral status is associated with both conditions? What respective effect do the first and second perspectives have on the degree of moral status? If we can only rescue either HBOs, which belong to the class of humans but do not have any psychological capabilities such as consciousness, or non-human beings such as robots having sophisticated consciousness, which should be rescued first? Furthermore, what exactly does occupying a moral status require us to morally refrain from doing, and in the case of HBOs, what are the ethical demands

19 As Sawai *et al*. correctly point out, the study of HBOs can be divided into at least three phases: *in vitro* studies of HBOs, transplantation of HBOs into animal brains, and transplantation of HBOs into human brains (Sawai *et al*. 2021). From this perspective, it might be fair to say that I have primarily examined only the first phase in this paper. Note that Sawai *et al*. also make a persuasive proposal in that they urge a focus on the precautionary principle prior to this paper, stating that "it may be advisable to adopt a precautionary approach that assumes future brain organoids will possess particular kinds of consciousness" (ibid.: 90).

on the scientists in the lab? There are many remaining questions, and none of which will be easy to answer. However, focusing not only on the possibility that HBOs can develop various kinds of "consciousness" individually, but also on the new possibility that HBOs are members of the "we" in terms of family membership, as are PVSs, is an important guide to seriously considering the moral status of HBOs. I end this paper with the hope that the abstract discussion engaged in here will serve as a springboard for a more detailed analysis of these points in the future.[20]

References

Block, N. (1995), 'On a Confusion about a Function of Consciousness', *Behavioral and Brain Sciences* 18 (2): 227–47.

Bollinger, J. *et al*. (2021), 'Patients' Perspectives on the Derivation and Use of Organoids', *Stem Cell Reports* 16 (8): 1874–83.

Bortolotti, L. and Harris, J. (2006), 'Embryos and Eagles: Symbolic Value in Research and Reproduction', *Cambridge Quarterly of Healthcare Ethics* 15 (1): 22–34.

Cordeiro-Rodrigues, L. and Ewuoso, C. (2021) 'An Afro-Communitarian Relational Approach to Brain Surrogates Research', *Neuroethics* 14 (3): 561–574.

DeGrazia, D. (2008) 'Moral Status as a Matter of Degree?', *Southern Journal of Philosophy* 46 (2): 181–98.

DeGrazia, D. (2021), 'An Interest-based Model of Moral Status', in S. Clarke, H. Zohny, and J. Savulescu (eds.), *Rethinking Moral Status*, New York: Oxford University Press, 40–56.

Eiraku, M. *et al*. (2008), 'Self-Organized Formation of Polarized Cortical Tissues from ESCs and its Active Manipulation by Extrinsic Signals', *Cell Stem Cell* 3 (5): 519–32.

Greely, H. T. (2021), 'Human Brain Surrogates Research: The Onrushing Ethical Dilemma', *American Journal of Bioethics* 21 (1): 34–45.

Kuhse, H. and Singer, P. (1990), 'Individuals, Humans, and Persons', in P. Singer, H. Kuhse, S. Buckle, K. Dawson, P. Kasimba (eds.), *Embryo Experimentation*, Cambridge: Cambridge University Press, 65–75.

Harman, E. (2007), 'Sacred Mountains and Beloved Fetuses', *Philosophical Studies* 133 (1): 55–81.

Hopster, J. (2019), 'The Speciesism Debate', *Animals* 9, 1054.

Hostiuc, S. *et al*. (2019), 'The Moral Status of Cerebral Organoids', *Regenerative Therapy* 10: 118–22.

Hyun, I. *et al*. (2020), 'Ethical Issues Related to Brain Organoid Research', *Brain Research* 1732, https://doi.org/10.1016/j.brainres.2020.146653.

International Society for Stem Cell Research (2021), 'ISSCR Guidelines for Stem Cell Research and Clinical Translation', 2021 update, https://www.isscr.org/policy/guidelines-for-stem-cell-research-and-clinical-translation/key-topics/organoids.

Kahane, G. (2009), 'Pain, Dislike and Experience', *Utilitas* 21 (3): 327–36.

Kahane, G. and Savulescu, J. (2009), 'Brain Damage and the Moral Significance of Consciousness', *Journal of Medicine and Philosophy* 34 (1): 6–26.

Kamm, F. M. (2007), *Intricate Ethics*, New York: Oxford University Press.

Kittay, E. F. (2005), 'At the Margins of Moral Personhood', *Ethics* 116 (1): 100–31.

Koplin, J. J. and Savulescu, J. (2019), 'Moral Limits of Brain Organoid Research', *Journal of Law, Medicine & Ethics* 47 (4): 760–67.

Kriegel, U. (2017), 'Dignity and the Phenomenology of Recognition-respect', in J. J. Drummond and S. Rinofner-Kreidl (eds.), *Emotional Experiences*, New York: Rowman & Littlefield, 121–36.

Lancaster, M. A. *et al*., (2013), 'Cerebral Organoids Model Human Brain Development and Microcephaly', *Nature* 501: 373–79.

Lavazza, A. and Pizzetti, F. G. (2020), 'Human Cerebral Organoids as a New Legal and Ethical Challenge', *Journal of Law and the Biosciences* 7 (1), https://doi.org/10.1093/jlb/lsaa005.

Lee, A. Y. (2019), 'Is Consciousness Intrinsically Valuable?', *Philosophical Studies* 176 (3), 655–71.

Levy, N. (2014), 'The Value of Consciousness', *Journal of Consciousness Studies* 21 (1–2): 127–38.

Levy, N. and Savulescu, J. (2009), 'Moral Significance of Phenomenal Consciousness', *Progress in Brain Research* 177: 361–70.

Lewis, D. (1986), *On the Plurality of Worlds*, Oxford: Blackwell.

Mackie, D. (1999), 'Personal Identity and Dead People', *Philosophical Studies* 95 (3): 219–242.

Mansour, A. A. *et al*., (2018), 'An i*n vivo* Model of Functional and Vascularized Human Brain Organoids', *Nature Biotechnology* 36: 432–41.

Matsui, T. K. *et al*. (2018), 'Six-month Cultured Cerebral Organoids from Human ES Cells Contain Matured Neural Cells', *Neuroscience Letters* 670 (23): 75–82.

Merricks, T. (2001), *Objects and Persons*, Oxford: Oxford University Press.

McMahan, J. (2002), *The Ethics of Killing*, New York: Oxford University Press.

Munsie, M. *et al*. (2017), 'Ethical Issues in Human Organoid and Gastruloid Research', *Development* 144 (6): 942–945

Niikawa, T. *et al*. (2022), 'Human Brain Organoids and Consciousness', *Neuroethics* 15, https://doi.org/10.1007/s12152-022-09483-1

Olson, E. T. (1997), *The Human Animal*, New York: Oxford University Press.

Parfit, D. (2008), 'Persons, Bodies, and Human Beings', in T. Sider, J. Hawthorne, and D. W. Zimmerman eds., *Contemporary Debates in Metaphysics*, Malden, MA: Blackwell.

Qian, X. *et al*. (2016), 'Brain-region-specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure', *Cell* 165 (5): 1238–54.

20 I would like to thank two anonymous reviewers of the journal for their highly helpful comments and suggestions. This work was supported by JSPS KAKENHI Grant Number JP20J00631.

Quadrato, G. *et al*. (2017), 'Cell Diversity and Network Dynamics in Photosensitive Human Brain Organoids', *Nature* 545: 48–53.

Robertson, J. A. (1999), 'Ethics and Policy in Embryonic Stem Cell Research', *Kennedy Institute of Ethics Journal* 9 (2): 109–36.

Sawai, T. *et al*. (2021), 'Mapping the Ethical Issues of Brain Organoid Research and Application', *AJOB Neuroscience* 13 (2): 81–94.

Scanlon, T. M. (1998), *What We Owe to Each Other*, Cambridge, MA: Belknap Press.

Sebo, J. (2018), 'The Moral Problem of Other Minds', *Harvard Review of Philosophy* 25: 51–70.

Singer, P. (1979), 'Killing Humans and Killing Animals', *Inquiry* 22 (1–4), 145–56.

Singer, P. (2009), *Animal Liberation*, updated edition, New York: Harper Perennial Modern Classics.

Snowdon, P. (2014), *Persons, Animals, Ourselves*, New York: Oxford University Press.

Trujillo, C. A. *et al*. (2019), 'Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development', *Cell Stem Cell* 25 (4): 558–69.

van Inwagen, P., (1990), *Material Beings*, Ithaca, NY: Cornell University Press.

Warren, M. A. (1997), *Moral Status*, New York: Oxford University Press.

Williams, B. (2006), 'The Human Prejudice', in his *Philosophy as a Humanistic Discipline*, A. W. Moore ed., Princeton, NJ: Princeton University Press.

A Four-dimensionalist Theory of Actions and Agents

Yasuo Nakayama
Emeritus professor, Graduate School of Human Sciences, Osaka University

Abstract

Donald Davidson proposed an event ontology and considered events as entities to which people can refer and over which people can quantify. Davidson also proposed to interpret actions as events that are intentional under some descriptions. Recent years, some philosophers further developed his theory of action and extended their analysis to collective actions. It is an aim of this paper to contribute to this research program.

In this paper, based on a four-dimensional event ontology, I propose an *axiomatic theory for actions and* agents and analyze collective actions and extended agents. This paper mainly investigates individual and collective actions that involve tools. This investigation also aims to give a metaphysical characterization of the ontological status of tools. In this investigation, it turns out that the four-dimensional ontology is a powerful framework to semantically describe temporally extended actions and collective actions. In two appendixes, I give a precise description of the proposed theories.

Keywords: collective action, extended agent, group agency, four-dimensional event ontology, ontology of artifacts

Introduction

Donald Davidson developed an event ontology and considered events as First-Order entities (FO-entities) as well as things. This means that we can quantify over events. Furthermore, he interpreted actions as events that are intentional under some descriptions (Davidson 1980). Since then, this event ontology of Davidson is widely accepted. For example, Kirk Ludwig further developed Davidson's event ontology and proposed a truth-conditional event ontology (Ludwig 2016; 2017).

In this paper, I extend the notion of *agent*. Some actions are performed by plural agents who use some tools. A play of symphony by an orchestra is a typical example for such actions. Recently, some philosophers started to investigate collective actions (Searle 2010; Tuomela 2013; Bratman 2014; Gilbert 2014; Ludwig 2016; 2017). However, actions performed with tools have been rarely studied, and one of topics in this paper is devoted for this problem. I propose to characterize a tool as a part of an extended agent. For this purpose, I use four-dimensional mereology as the formal framework. Nakayama (2013) proposed a notion of *extended agent*. However, this characterization of *extended agent* was not sufficient because notion of *joint action* (i.e., *collective action*) that is a key concept for this study was insufficiently characterized. To overcome this problem, this paper aims to clarify notions of *extended agent* and *collective action*.

There are several formal devices to represent collective expressions. One framework is Plural Logic[1] which is used in semantic investigations by Ludwig (2016; 2017). Another framework is mereology. To develop an event ontology, Nakayama (2017) proposed to give a four-dimensional interpretation of mereological *parthood* and developed a four-dimensional event ontology.

In this paper, I propose an *axiomatic theory for actions and agents* that formally characterizes actions and agents. In the main text, I explain the framework and its application examples. In two appendixes, I give a precise formalism of the theory.

1. Four-dimensional Event Ontology

Ontological studies tend to focus on things. By contrast, Nakayama (2017; 2019) proposed a four-dimensional mereological system and developed

1 For Plural Logic, see Linnebo (2022).

an ontological framework that considers events (or processes) as fundamental entities. This formal system is called *four-dimensional event ontology* (Th(4EO)) or *process ontology*. Th(4EO) is a theory for events and interprets things as a kind of events. In this framework, the universe is defined as the maximal event and all other concrete objects are considered as parts of the universe (see (Ap1.2.s)). In this section, I explain an essential part of Th(4EO) that is useful in this paper.

A mereology is an axiomatic system that has *part* as the single primitive relation. Core Mereology is the theory of partial ordering (see Ap1.1.a+b+c), and the standard mereological system is General Extensional Mereology (GEM)[2]. Core mereology is the weakest system and GEM has a strong expressive power. There are many mereological systems between them and above them. From GEM follows the following two important principles.

(S1.1.a) [Extensionality] If x and y have the same proper parts, then $x = y$.
(S1.1.b) [Unrestricted composition] Every plurality of objects possesses a sum (i.e., fusion).

Thus, to accept GEM means to accept all composed entities as existing. Some philosophers are against this inflation of entities and prefer a mereological system that is weaker than GEM. There are many criticisms on GEM, but some problems can be solved by taking four-dimensionalsm (see Sider 2001). In this paper, GEM is used as the fundamental ontological theory and the event parthood is four-dimensionally interpreted. In fact, the acceptance of GEM does not change our world but only requires a shift of our view on existence. If you accept GEM, then you say not only objects A and B but also object A+B exist.[3] Note that A+B does not add any new (atomic) entity into the world, but it only adds new referents. By accepting GEM, we can refer to more objects than before, and this gives GEM a strong expressive power.

A four-dimensionalist interprets the parthood relation as a relation between 4D (four-dimensional) entities.[4] In this paper, GEM is accepted as the basic system and the event parthood is four-dimensionally interpreted. In this section, I provide the four-dimensional mereology for events that gives a basis for analysis of actions and agents. In four-dimensionalism, the notion of *temporal part* plays an essential role, where there are different versions of definition of this notion (see Hawley 2020). Here, I modify the definition in Nakayama (2017) (see (S1.2.h+i)). Now, I explain an outline of *four-dimensional event ontology*.

(S1.2) Core of Th(4EO)[5]
(a)[(Ap1.2.a+b+c)] I use three parthood relations, namely *part*, $part_{st}$, and $part_T$ (see Table 1). I assume GEM for *part*, GEM for $part_{st}$, and GEM for $part_T$.

Parthood relation	Interpretation
part	four-dimensional parthood relation for entities in the universe
$part_{st}$	four-dimensional parthood relation for space-time objects
$part_T$	one-dimensional parthood relation for time objects

Table 1. Three parthood relations

(b) [(Ap1.2.d)] Every event $occupies_{st}$ exactly one space-time region.
(c) [(Ap1.2.e)] Every event $occupies_T$ exactly one time region.
(d) [(Ap1.2.f)] S is the space-time region of E iff E $occupies_{st}$ S. The space-time region of E is expressed by $st(E)$.
(e) [(Ap1.2.g)] T is the existence time of E iff E $occupies_T$ T. The existence time of E is expressed by *exist-time*(E).
(f) [(Ap1.2.h)] The event parthood determines the space-time parthood. In other words, if E_1 is a *part* of E_2, then $st(E_1)$ is a $part_{st}$ of $st(E_2)$.
(g) [(Ap1.2.i)] The event parthood determines the temporal parthood. In other words, if E_1 is a *part* of E_2, then *exist-time*(E_1) is a $part_T$ of *exist-time*(E_2).
(h) [(Ap1.2.j)] E_1 is a *temporal part* of E_2 iff for every event E_3 [if *exist-time*(E_3) = *exist-time*(E_1), then [E_3 is a *part* of E_1 iff E_3 is a *part* of E_2]]. In short, E_1 is a *temporal part* of E_2 iff E_1 and E_2 are indistinguishable within the existence time of E_1.
(i) [(Ap1.2.k)] If T is a $part_T$ of *exist-time*(E_1), then [the temporal part of E_1 at $T = E_2$ iff [E_2 is a temporal part of E_1 and T = *exist-time*(E_2)]]. The temporal part of E at T is expressed by *temporal-part*(E, T).
(j) [(Ap1.2.m)] x is a $part_{tp}$ of y in E iff *temporal-part*(x, *exist-time*(E)) is a part of *temporal-part*(y, *exist-time*(E)).
(k) [(Ap1.2.n] E_1 is a *spatial part* of E_2 iff [E_1 is a *part* of E_2 and *exist-time*(E_1) = *exist-time*(E_2)].
(l) [(Ap1.2.r)] $T_1 < T_2$ iff the latest time point of T_1 is

2 For mereology and GEM, see Varzi (2019). GEM is an axiomatic theory in First-Order Logic (FO-Logic) and is decidable (Tsai 2013; Varzi 2019, Sect 4.4). For formalization of GEM, see (Ap1.1).
3 Here, + is used as the symbol for *mereological sum* (see (Ap1.1.i)).
4 In general, a four-dimensionalist needs not accept GEM. In fact, there are many four-dimensionalists who reject GEM (see Sider 2001).
5 The related statement in appendixes is indicated through caption in form [(Ap*m.n.α*)].

earlier than the earliest time point of T_2.
(m) [(Ap1.2.s)] The universe is the maximal event.
(n) [(Ap1.2.t)] *now* is an indexical that denotes the current time point.

Th(4EO) delivers a basis for the following discussions in this paper.

2. Actions and Atomic Agents

Philosophy of action started with studies on simple actions that are performed by atomic agents. In this context, Davidson proposed to characterize agency in terms of intention: "a man is the agent of an act if what he does can be described under an aspect that makes it intentional" (Davidson 1980, 46). This thesis has been quite influential in philosophy of action, and many philosophers applied this principle not only to simple actions but also to collective actions (Ludwig 2017, Chapter 2). However, it is difficult to describe what collective intention is. At least, the existence of collective intention is not so obvious as individual intention. In this paper, I characterize actions as events that are brought about by agents, and I do not presuppose that there is unique explanation for agency. I only try to give some semantic characterizations of agency.

The second fundamental observation of this paper is that events and actions are, in general, *temporally extended*. For example, a walk has a duration and exists for certain time. In the same way, an agent who performs an action has a temporal duration. Only for the time span of a walk, *the agent of this walk* exists. I interpret this temporary extendedness of actions based on Th(4EO) and propose the following ontological theses of the four-dimensional action theory.

(S2.1) Characterizations of four-dimensional action theory
(a) [(Ap1.2.s)] The *universe* is the maximal 4D-entity. This means that any 4D-entity is a part of the universe.
(b) An *event* is a 4D-entity. Thus, an action is also a 4D-entity.
(c) An *agent* is a 4D-entity.
(d) [(Ap1.3.a)] x is an $agent_{tp}$ of E iff *temporal-part*(x, *exist-time*(E)) is an agent of E.
(e) [(Ap1.3.d)] E is an action iff there is x such that x is an $agent_{tp}$ of E.

Here, term *agent* is used as a primitive notion and *action* is defined based on this notion. Definition (S2.1.e) means that an action is an event that has an $agent_{tp}$. In other words, an $agent_{tp}$ is an entity that produces an event, and such an event can be interpreted as an action.

Some actions have their objects. For example, a *throw of a ball* has a ball as an object of this action. I propose the following characterization of $object_{tp}$.

(S2.2.a) [(Ap1.3.b)] x is an $object_{tp}$ of E iff *temporal-part*(x, *exist-time*(E)) is an object of E.
(S2.2.b) [(Ap1.3.j)] Any $object_{tp}$ of E has an $agent_{tp}$ of E and any musical composition has an $agent_{tp}$ of E, when it is an object of E.

According to (S2.2.b), there is no $object_{tp}$ of E without an $agent_{tp}$ of E. Thus, any $object_{tp}$ of an action requires its $agent_{tp}$. This requirement is reasonable because an action always presupposes its $agent_{tp}$ (see (S2.1.e)).

An *atomic* $agent_{tp}$ is a simple agent and this notion is a key concept of the agency.

(S2.3) Characterizations of *atomic* $agent_{tp}$
(a) [(Ap1.3.f)] An *atomic* $agent_{tp}$ of E is an $agent_{tp}$ of E.
(b) [(Ap1.3.g)] If x is an *atomic* $agent_{tp}$ of E_1, E_2 is a *spatial part* of E_1, y is an $agent_{tp}$ of E_2, and y is a *part* of x, then $y = x$ and $E_2 = E_1$. This means that an *atomic* $agent_{tp}$ of E_1 has no member who performs a sub-action of E_1. In other words, an *atomic* $agent_{tp}$ of E_1 is the single indivisible $agent_{tp}$ who performs E_1.
(c) [(Ap1.3.h)] Any $agent_{tp}$ x of E_1 has an *atomic* $agent_{tp}$ y of E_2 as its part so that E_2 is a part of E_1 and y is a $part_{tp}$ of x in E_2. In short, any $agent_{tp}$ of an action has an *atomic* $agent_{tp}$ of its sub-action.

According to (S2.3), an *atomic* $agent_{tp}$ is the smallest $agent_{tp}$ who can be a constituent of other complex $agent_{tp}$. Furthermore, any $agent_{tp}$ is herself an *atomic* $agent_{tp}$ or has an *atomic* $agent_{tp}$ as a part.

By using the introduced notions, we can analyze some simple English sentences (We use *tr* as a translation function from English sentences into sentences in FO-Logic).

(S2.4) Some English sentences
(a) Let core[S2.4.a](E) be an abbreviation of (*atomic-*$agent_{tp}$(Mary, E) $\wedge$ singing(E) $\wedge$ *exist-time*(E) $<$ *now*).
(b) *tr*(Mary was singing): $\exists E$ core[S2.4.a](E).
(c) core[S2.4.c](E, a): (core[S2.4.a](E) $\wedge$ *object*(a, E) $\wedge$ song(a)).
(d) *tr*(Mary sang a song): $\exists E\, \exists a$ core[S2.4.c](E, a).[6]

6 A musical composition is an abstract object. For quantifications over musical compositions, see (Ap1.2.a+b). A set of abstract objects is often accepted as a domain in FO-Logic. For example, numbers and many mathematical objects are abstract objects that can be described in FO-Logic. However, I do not discuss this problem of abstract objects in this paper.

(e) core[S2.4.e](E, x, a): (core[S2.4.c](E, a) ∧ in($st(E)$, $st(x)$) ∧ school(x)).
(f) tr(Mary sang a song in a school): $\exists E\ \exists x\ \exists a$ core[S2.4.e](E, x, a).
(g) tr(Mary sang a song in a school yesterday): $\exists E\ \exists x\ \exists a$ (core[S2.4.e](E, x, a) ∧ $part_T$(*exist-time*(E), yesterday(*now*))).
(h) tr(Someone performs an action with a song): $\exists E\ \exists x\ \exists a$ (*agent*(x, E) ∧ *action*(E) ∧ $object_{tp}$(a, E) ∧ song(a)).
(i) In FO-Logic, the following sentences are valid.
  (i1) tr(Mary sang a song) → tr(Mary was singing)
  (i2) tr(Mary sang a song in a school) → tr(Mary sang a song)
  (i3) tr(Mary sang a song in a school yesterday) → tr(Mary sang a song in a school)
  (i4) tr(Mary sang a song) → $\exists E\ \exists a$ (*object*(a, E) ∧ song(a))
(j) Based on (3.2.1.e) and (3.2.2.b), from Th(AT) follows: $\exists E\ \exists a$ ((*object*(a, E) ∧ song(a)) → tr(Someone performs an action with a song)).

Davidson suggested that inferences in (S2.4.i1+i2+i3) are provable in his framework (Davidson 1967). According to (S2.4.i4), it appears that an action without an agent is possible. However, Th(AT) blocks this consequence because any object of an action implies the existence of an agent according to (S2.2.b).

English speakers share a belief base, and I claim that Th(AT) is included in this shared belief base.

3. Actions and Extended Agents

In this section, I would like to clarify some features of tools. Tools are artifacts that are used to support actions. To properly use a tool, we often need some exercises so that the tool is properly integrated into actions when we perform with it. Based on this observation, I propose (S3.1).

(S3.1) Characterizations of tools
(a) [(Ap1.3.c)] x is a $tool_{tp}$ for E iff *temporal-part*(x, *exist-time*(E)) is a tool for E.
(b) [(Ap1.3.k)] Any $tool_{tp}$ x for E has an $agent_{tp}$ of E which includes x as a part in E. In short, any $tool_{tp}$ x for E has an $agent_{tp}$ of E in which x is a constituent of this $agent_{tp}$ of E.
(c) [(Ap1.3.l)] If x+y is an $agent_{tp}$ of E_1 and y is a $tool_{tp}$ for E_1, then there is x's action E_2 such that E_2 is a *spatial part* of E_1 and x uses y as an $object_{tp}$ of E_2.
(d) [(Ap1.3.i)] An *extended* $agent_{tp}$ of E is an $agent_{tp}$ of E who is no *atomic* $agent_{tp}$ of E.

(S3.1.b) expresses a fundamental feature of tools. An entity becomes a tool only for the time in which an $agent_{tp}$ uses it to perform an action. For example, a knife is a simple entity when nobody uses it. Only when someone uses it, it functions as a tool.

According to (S3.1.d) and (S2.3.b), $agent_{tp}$ x+y of an action is an extended $agent_{tp}$ of this action when x+y is constructed from $agent_{tp}$ x and a $tool_{tp}$ y. A tool expands the power of an agent. A nearsighted person can safely drive a car when she wears glasses. People can easily communicate with each other in long distance when they have smart phones. As often mentioned, it is an essential capacity of humans to invent tools and to use them.

In natural languages, the existence of tools is often not explicitly expressed. In such cases, I propose to supplement a subject of a sentence with expression *with a tool*. For example, let us think about sentence *John buttered a piece of toast*. Obviously, John cannot butter a piece of toast without tool, so that I interpret the sentence as *John (with a tool) buttered a piece of toast*.

(S3.2.a) John-buttered: John (with a tool) buttered a piece of toast.
(S3.2.b) core[S3.2.b](E, x, y): ($agent_{tp}$(John+x, E) ∧ buttering(E) ∧ $tool_{tp}$(x, E) ∧ $object_{tp}$(y, E) ∧ toast(y) ∧ *exist-time*(E) < *now*).
(S3.2.c) tr(John-buttered): $\exists E\ \exists x\ \exists y$ core[S3.2.b](E, x, y).
(S3.2.d): tr(John-buttered, where John used the tool): $\exists E_1\ \exists x\ \exists y$ (core[S3.2.b](E_1, x, y) ∧ $\exists E_2$ ($agent_{tp}$(John, E_2) ∧ use(E_2) ∧ *sp-part*(E_2, E_1) ∧ $object_{tp}$(x, E_2))).

Here, *John (with a tool)* is an extended agent. Because of (S3.1.c), there is a sub-action of John that he performed using the tool.

(S3.2.e) Based on (S3.1.c), from Th(AT) follows: tr(John-buttered) → tr(John-buttered, where John used the tool).

Level 1	John (with a tool) buttered a piece of toast.	
Level 2	John used the tool.	

Table 2. Use of a tool

Here, a tool appears in two different modes, namely in the mode of a part of an extended agent and in the mode of an object of the atomic agent who is a part of the extended agent. John has a leading desire to butter a piece of toast with an object and John intends to move the object so that this leading desire will be satisfied. This John's action can be described as two layers of descriptions (Table 2). When John butters a piece of toast, he moves a tool so that this movement of the tool

realizes the buttering a piece of toast. In this case, there are two actions, but the action on Level 2 constitutes an important part of the action of the extended agent on Level 1.

In this paper, I interpret sentence *John buttered a piece of toast with a knife* as *John (with a tool) buttered a piece of toast and this tool is a knife.*

(S3.3.a) *tr*(John buttered a piece of toast with a knife): $\exists E\ \exists x\ \exists y$ (core[S3.2.b](E, x, y) $\wedge$ knife(x)).
(S3.3.b) Based on (S3.2.c) and (S3.3.a), it is valid in FO-Logic: *tr*(John buttered a piece of toast with a knife) $\rightarrow$ *tr*(John (with a tool) buttered a piece of toast).

There are cases in which extended agents appear quite natural. Suppose that Paul lost his left leg by an accident, and since then he usually uses an artificial leg. When Paul is walking equipped with the artificial left leg, it is appropriate to say, "Paul with his artificial left leg is walking". Or, when we say, "Paul is walking", we mean that Paul with his artificial left leg is walking. Here, the agent of walking should be interpreted as *Paul with his artificial left leg*.

Now, let us consider a sentence in which two tools appear: "Booth shot Lincoln with a gun and pulled the trigger with his finger". It is known that the event semantics raises a problem when the shooting action and the pulling action are the same action under different descriptions (Ludwig 2018, 477). In our interpretation, the pulling action is a four-dimensional proper part of the shooting action so that the reported problem does not occur.

(S3.4) *tr*(Booth shot Lincoln with a gun and pulled the trigger with his finger): $\exists E_1\ \exists E_2\ \exists x\ \exists y\ \exists z$ (*agent*$_{tp}$(Booth+x, E_1) $\wedge$ shooting(E_1) $\wedge$ *tool*$_{tp}$(x, E_1) $\wedge$ gun(x) $\wedge$ *object*$_{tp}$(Lincoln, E_1) $\wedge$ *exist-time*(E_1) < *now* $\wedge$ *agent*$_{tp}$(Booth, E_2) $\wedge$ pulling(E_2) $\wedge$ *proper-part*(E_2, E_1) $\wedge$ *tool*$_{tp}$(y, E_2) $\wedge$ finger(y) $\wedge$ *part*(y, Booth) $\wedge$ *object*$_{tp}$(z, E_2) $\wedge$ trigger(z) $\wedge$ *part*(z, x)).

According to the translation (S3.4), the gun is a tool for shooting, Booth's finger is a tool for pulling the trigger of the pistol, and the action of *shooting the gun* includes the action of *pulling the trigger* as a proper part. To shoot Lincoln, Booth must exactly target him with the gun. Thus, shooting a gun is not just pulling a trigger. Here, we see that it is sometimes important to take four-dimensional extendedness of actions into consideration.

4. Collective Actions and Collective Agents

Ludwig (2017) pointed out that collective actions deal with the collective reading of plural and singular group action sentences (Chapter 3 and 4). To express the collective reading of plural and singular group action sentences in Th(AT), I introduce two notions, namely *group*$_{tp}$ and *member*$_{tp}$.

(S4.1) Characterizations of *group*$_{tp}$ and *member*$_{tp}$
(a) [(Ap1.3.m)] x is a *group*$_{tp}$ for E iff *temporal-part*(x, *exist-time*(E)) is a group for E.
(b) [(Ap1.3.n)] If x is a *group*$_{tp}$ for E and y is a *group*$_{tp}$ for E, then $x = y$.
(c) [(Ap1.3.o)] If x is a *group*$_{tp}$ for E_1, then [y is a *member*$_{tp}$ of x for E_1 iff [*temporal-part*(y, *exist-time*(E_1)) is a member of *temporal-part*(x, *exist-time*(E_1)), and there is E_2 such that E_2 is a proper part of E_1 and y is an *agents*$_{tp}$ of E_2]].
(d) [(Ap1.3.p)] At least two *members*$_{tp}$ for E belongs to a *group*$_{tp}$ for E.

I use *group for E* in the meaning of *group that is formed for the execution of E*. This idea is reflected in (S4.1.a). Corresponding to this idea, *members*$_{tp}$ of a *group*$_{tp}$ for E are stipulated as *agents*$_{tp}$ who contribute to a successful execution of E by performing sub-actions of E (see (S4.1.c)).

To explain collective actions, let us consider plural action sentence *They walked*. According to the distributing reading, the sentence means that each of them walked. By contrast, according to the collective reading, the sentence means that they walked together. In a language that includes Th(AT), these meanings can be expressed as follows.

(S4.2) Sentences with a walk
(a) core[S4.2.a](x, E): (*agent*$_{tp}$(x, E) $\wedge$ walking(E) $\wedge$ *exist-time*(E) < *now*).
(b) *tr*(Mary walked): $\exists E$ (core[S4.2.a](Mary, E) $\wedge$ *atomic-agent*$_{tp}$(Mary, E)).
(c) *tr**(They walked, distributive): $\forall x$ (*member*(x, They) $\rightarrow$ $\exists E$ core[S4.2.a](x, E)).
(d) *tr**(They walked, collective): $\exists E$ (core[S4.2.a](They, E) $\wedge$ *group*$_{tp}$(They, E)).
(e) *tr**(Mary and Tom walked, distributive): *tr*(Mary walked) $\wedge$ *tr*(Tom walked).
(f) core[S4.2.f](E): (G = Mary+Tom $\wedge$ core[S4.2.a](G, E) $\wedge$ *group*$_{tp}$(G, E)).
(g) *tr**(Mary and Tom walked, collective): $\exists E$ core[S4.2.f](E).
(h) *tr**(Mary and Tom walked together as its members, collective): $\exists E$ (core[S4.2.f](E) $\wedge$ *member*$_{tp}$(Mary, G, E) $\wedge$ *member*$_{tp}$(Tom, G, E)).

According to the distributive readings (S4.2.c+e), every member of them performed a walking action in the past. By contrast, in the collective readings (S4.2.d+g),

the action of each member is not explicitly mentioned. In a collective walking E, the $members_{tp}$ for E performed a sub-action of E so that the $group_{tp}$ for E could accomplish E. The sentence does not express what kind of actions the $members_{tp}$ for E performed.

Here, I propose to interpret a collective action as an action whose $agent_{tp}$ is a group of agents.

(S4.3) Characterizations for collective actions
(a) [(Ap1.3.q)] A *collective-action*$_{tp}$ is an action whose $agent_{tp}$ is a $group_{tp}$.
(b) [(Ap1.3.r)] If E is a *collective-action*$_{tp}$ and x is a $group_{tp}$ for E, then E consists of actions which are performed by the $agents_{tp}$ who are $members_{tp}$ of x for E.

Here, let us think again about the collective walk of Mary and Tom. In this case, because of (S4.3.b), from Th(AT) follows: *tr**(Mary and Tom walked together as its members, collective) → $\exists E_1\ \exists E_2\ \exists E_3$ ($agent_{tp}$(Mary+Tom, E_1) ∧ walking(E_1) ∧ $E_1 = E_2+E_3$ ∧ $agent_{tp}$(Mary, E_2) ∧ *action*(E_2) ∧ $agent_{tp}$(Tom, E_3) ∧ *action*(E_3)). Thus, when Mary and Tom are walking together, they try to perform individual actions that support the accomplishment of their collective walk. This individual action in a collective walk requires more than a simple individual walk and this further requirement distinguishes a collective walk from a simple fusion of plural individual walks (see Table 3 and Table 4).

Level 1	Mary and Tom walked together.	
Level 2	Mary performed a sub-action of the collective action in Level 1.	Tom performed a sub-action of the collective action in Level 1.

Table 3. Collective reading of *Mary and Tom walked*

Level 1	Mary walked.	Tom walked.

Table 4. Distributive reading of *Mary and Tom walked*

Next, let us consider a more complex example: A violinist and a pianist played Beethoven's *Spring Sonata* (BSS). In this example, a collective action of playing a sonata was performed by two extended agents who used music instruments.

(S4.4) Play of BSS
(a) core[S4.4.a](E, G, x, y): ($agent_{tp}$(G, E) ∧ $group_{tp}$(G, E) ∧ $part_{tp}$(x+y, G, E) ∧ violinist(x) ∧ pianist(y) ∧ playing(E) ∧ *object*(BSS, E) ∧ *exist-time*(E) < *now*).
(b) *tr**(A violinist and a pianist played BSS, collective): $\exists E\ \exists G\ \exists x\ \exists y$ core[S4.4.a](E, G, x, y).
(c) core[S4.4.c](E_1, E_2, E_3, G, x, y, u, v): (core[S4.4.a](E_1, G, x, y) ∧ $G = x+u+y+v$ ∧ $member_{tp}$(x+u, G, E_1) ∧ $member_{tp}$(y+v, G, E_1) ∧ $E_1 = E_2+E_3$ ∧ $agent_{tp}$(x+u, E_2) ∧ $tool_{tp}$(u, E_2) ∧ violin(u) ∧ $agent_{tp}$(y+v, E_3) ∧ $tool_{tp}$(v, E_3) ∧ piano(v)).
(d) *tr**(A violinist (with a violin) and a pianist (with a piano) played BSS, collective): $\exists E_1\ \exists E_2\ \exists E_3\ \exists G\ \exists x\ \exists y\ \exists u\ \exists v$ core[S4.4.c](E_1, E_2, E_3, G, x, y, u, v).
(e) core[S4.4.e](E_1, E_2, E_3, E_4, E_5, G, x, y, u, v): (core[S4.4.c](E_1, E_2, E_3, G, x, y, u, v) ∧ *sp-part*(E_4, E_2) ∧ $agent_{tp}$(x, E_4) ∧ use(E_4) ∧ $object_{tp}$(u, E_4) ∧ *sp-part*(E_5, E_2) ∧ $agent_{tp}$(y, E_5) ∧ use(E_5) ∧ $object_{tp}$(v, E_5)).
(f) Because of (S3.1.c), from Th(AT) follows: *tr**(A violinist (with a violin) and a pianist (with a piano) played BSS, collective) → $\exists E_1\ \exists E_2\ \exists E_3\ \exists E_4\ \exists E_5\ \exists G\ \exists x\ \exists y\ \exists u\ \exists v$ core[S4.4.e](E_1, E_2, E_3, E_4, E_5, G, x, y, u, v).

Level 1	A violinist and a pianist played BSS.			
Level 2	The violinist with a violin played a part of BSS as a sub-action of the collective action in Level 1.		The pianist with a piano played a part of BSS as a sub-action of the collective action in Level 1.	
Level 3	The violinist used the violin.			The pianist used the piano.

Table 5. Play of BSS

As this example shows, a collective action can have a quite complex structure. In the play of BSS, the violinist moves a string of the violin to play her part of BSS, and the pianist touches keys of the piano to play her part of BSS. Their plays of instruments must be coordinated so that the sum of both activities produces BSS. Otherwise, they cannot successfully play the sonata (see Table 5).

In some collective actions, different types of division of labor are possible. Let us consider a collective painting. Suppose that Peter and Tom decide to paint a house. They might divide the task of painting so that Peter (with a tool) paint a part of the house and Tom (with a tool) paint another part (see (S4.5.d+e)). In other situation, it might be the case that only Peter is a skillful painter and Tom devotes himself to support of Peter's painting (see (S4.5.f+g)).

(S4.5) Collective actions with different divisions of labor
(a)Sentence-painting: Peter and Tom painted a house.
(b) core[S4.5.b](E, x, y): (G(x) = Peter+Tom+x ∧ $agent_{tp}$(G(x), E) ∧ $group_{tp}$(G(x), E) ∧ $tool_{tp}$(x, E) ∧ painting(E) ∧ $object_{tp}$(y, E) ∧ house(y) ∧ *exist-time*(E) < *now*).
(c) *tr**(Sentence-painting, collective): $\exists E\ \exists x\ \exists y$ core[S4.5.b](E, x, y).
(d) core[S4.5.d](E_1, E_2, E_3, x, y, u, v, $y1$, $y2$):

(core[S4.5.b])(E_1, x, y) ∧ $E_1 = E_2$+E_3 ∧ $agent_{tp}$(Peter+u, E_2) ∧ $member_{tp}$(Peter+u, G(x), E_1) ∧ $tool_{tp}$(u, E_2) ∧ $x = u$+v ∧ painting(E_2) ∧ $object_{tp}$($y1$, E_2) ∧ $y = y1$+$y2$ ∧ $agent_{tp}$(Tom+v, E_2) ∧ $member_{tp}$(Tom+v, G(x), E_1) ∧ $tool_{tp}$(v, E_3) ∧ painting(E_3) ∧ $object_{tp}$($y2$, E_3)).

(e) *tr**(Sentence-painting and they divided the task, collective): ∃E_1 ∃E_2 ∃E_3 ∃x ∃y ∃u ∃v ∃$y1$ ∃$y2$ core[S4.5.d](E_1, E_2, E_3, x, y, u, v, $y1$, $y2$).

(f) core[S4.5.f](E_1, E_2, E_3, x, y): (core[S4.5.b](E_1, x, y) ∧ $E_1 = E_2$+E_3 ∧ $member_{tp}$(Peter+x, G(x), E_1) ∧ $agent_{tp}$(Peter+x, E_2) ∧ painting(E_2) ∧ $object_{tp}$(y, E_2) ∧ $member_{tp}$(Tom, G(x), E_1) ∧ $agent_{tp}$(Tom, E_3) ∧ supporting(E_3) ∧ $object_{tp}$(Peter+x, E_3)).

(g) *tr**(Sentence-painting and Tom supported Peter during the work, collective): ∃E_1 ∃E_2 ∃E_3 ∃x ∃y core[S4.5.f](E_1, E_2, E_3, x, y).

This example shows that there are different forms of cooperation in collective actions. In some cases, tasks in a collective action are equally distributed among members of the collective agent. On the other hand, there are cases in which some agents perform main actions and other members support them.

There are collective actions that have a layered structure. Let us consider the following example of production in a factory.

(S4.6) Collective action with a layered structure of actions

(a) Example: There was a factory F_1 and there were two machines M_1 and M_2 in F_1. Workers of F_1 produced packed flue masks with the machines. Group G_1 of workers produced flue masks with machine M_1 and group G_2 of workers packed the products with machine M_2.

(b) core[S4.6.b]: (factory(F_1) ∧ machine(M_1) ∧ machine(M_2) ∧ M = M_1+ M_2 ∧ in(*st*(M), *st*(F_1)) ∧ *exist-time*(F_1) < *now*).

(c) [The first sentence] *tr*(1): core[S4.6.b].

(d) core[S4.6.d](E, x): (core[S4.6.b] ∧ G = σu worker-of(u, F_1) ∧ $agent_{tp}$(G+M, E) ∧ $group_{tp}$(G+M, E) ∧ producing(E) ∧ $tool_{tp}$(M, E) ∧ $object_{tp}$(x, E) ∧ packed-flue-masks(x) ∧ *exist-time*(E) < *now*).[7]

(e) [The first two sentences] *tr*(1+2): ∃E ∃x core[S4.6.d](E, x).

(f) core[S4.6.f](E_1, E_2, E_3, x, y): (core[S4.6.d](E_1, x) ∧ $E_1 = E_2$+E_3 ∧ G = G_1+G_2 ∧ $agent_{tp}$(G_1+M_1, E_2) ∧ $group_{tp}$(G_1+M_1, E_2) ∧ producing(E_2) ∧ $object_{tp}$(y, E_2) ∧ flue-masks(y) ∧ *part*(y, x) ∧ $agent_{tp}$(G_2+M_2, E_3) ∧ $group_{tp}$(G_2+M_2, E_3) ∧ packing(E_3) ∧ $object_{tp}$(x, E_3)).

(g) [Three sentences] *tr*(1+2+3): ∃E_1 ∃E_2 ∃E_3 ∃x ∃y core[S4.6.f](E_1, E_2, E_3, x, y).

(h) Based on (S4.3.b), from Th(AT) follows: *tr*(1+2+3) → ∃E_1 ∃E_2 ∃E_3 ∃x ∃y (core[S4.6.f](E_1, E_2, E_3, x, y) ∧ ∀z ($member_{tp}$(z, G_1+M_1, E_2) → ∃E_4 (*part*(E_4, E_2) ∧ $agent_{tp}$(z, E_4))) ∧ ∀z ($member_{tp}$(z, G_2+M_2, E_3) → ∃E_4 (*part*(E_4, E_3) ∧ $agent_{tp}$(z, E_4)))).

Level 1	G produced pfms (packed flue masks) with M.			
Level 2	G_1 produced fms with M_1 as a sub-action of the collective action in Level1.		G_2 packed pfms with M_2 as a sub-action of the collective action in Level1.	
Level 3	Workers in G_1 with M_1 performed sub-actions of the collective action in Level 2.		Workers in G_2 with M_2 performed sub-actions of the collective action in Level 2.	
Level 4	Each worker in G_1 performed an action with a part of M_1 as a part of action described in Level 3.			Each worker in G_2 performed an action with a part of M_2 as a part of action described in Level 3.

Table 6. Production with two machines

In this example, the production of packed flue masks is divided into two parts, namely production of flue masks and packing of them. This is *division of labor* whose notion was introduced by sociologist Émil Durkheim (Durkheim 1893). In this case, the form of the division was determined by the features of two machines. The structure of machines restricts the range of possible distributions of workers. In the example, we can identify four levels of working stages (see Table 6). As Table 6 shows, a collective action can have a complex layered structure. Note that we can construct more complex layered structure of collective actions.

Based on Th(AT), we can properly express the content of a sentence which is a conjunction of a plural subject action sentence with the collective reading and one with the distributive reading.

(S4.7) Collective reading and distributive reading

(a)Example[S4.7] (Ludwig (2016), 143): They carried the piano upstairs and got a cookie as a reward.

(b) core[S4.7.b](E, x): ($agent_{tp}$(They, E) ∧ $group_{tp}$(They, E) ∧ carrying-upstairs(E) ∧ $object_{tp}$(x, E) ∧ piano(x) ∧ *exist-time*(E) < *now*).

(c) *tr**(They carried the piano upstairs, collective): ∃E ∃$^{=1}x$ core[S4.7.b](E, x).

(d) core[S4.7.d](E_1, E_2, x): (core[S4.7.b](E_1, x) ∧ ∀y (*member*(y, *temporal-part*(They, E_2)) → ∃E_3 ∃z (get-as-a-reward(*temporal-part*(y, E_3), E_3) ∧ *part*(E_3, E_2) ∧ $object_{tp}$(z, E_3) ∧ cookie(z) ∧ *exist-time*(E_1) < *exist-time*(E_3) ∧ *exist-time*(E_3) < *now*))).

7 σu worker-of(u, F_1) denotes the sum of workers of factory F_1. For this notation, see (Ap1.1.g).

(e) tr^*(Example[S4.7], collective): $\exists E_1\ \exists\ E_2\ \exists x$ core[S4.7.d](E_1, E_2, x).

In this example, the plural subject "they" in the collective reading denotes a temporal part of a plural entity which is a $group_{tp}$ for a carrying action and the second plural subject in the distributive reading denotes a different temporal part of the same plural entity.

The examples in this section suggest that Th(AT) can be used to analyze complex human activities.

5. Characterizations of Actions

As a summary of this paper, I describe translation schemata for sentences with transitive verbs. Translation schemata for sentences with intransitive verbs can be constructed in a similar manner (see (S4.2)).

(S5.1) Translation schemata for sentences with transitive verbs

(a) core[S5.1.a](E, x, y): ($agent_{tp}(x, E) \wedge \text{act}_{\text{transitive}}(E) \wedge object_{tp}(y, E) \wedge \text{obj}(y) \wedge exist\text{-}time(E) < now$).

(b) [Singular sentence] tr(S [$\text{act}_{\text{transitive}}$]$_{\text{past}}$ an obj): $\exists E\ \exists x$ (core[S5.1.a] (E, S, x) $\wedge$ $atomic\text{-}agent_{tp}$(S, E)).

(c) [Singular sentence + tool] tr(S [$\text{act}_{\text{transitive}}$]$_{\text{past}}$ an obj with a device): $\exists E\ \exists x\ \exists y$ (core[S5.1.a](E, S+y, x) $\wedge$ $atomic\text{-}agent_{tp}$(S, E) $\wedge$ $tool_{tp}$(y, E) $\wedge$ device(y)).

(d) [Plural sentence] tr^*(G [$\text{act}_{\text{transitive}}$]$_{\text{past}}$ an obj, distributive): $\forall x\ (member(x, \text{G}) \rightarrow \exists E\ \exists y$ core[S5.1.a](E, x, y)).

(e) [Plural sentence] tr^*(G [$\text{act}_{\text{transitive}}$]$_{\text{past}}$ an obj, collective): $\exists E\ \exists x$ (core[S5.1.a](E, G, x) $\wedge$ $group_{tp}$(G, E)).

(f) [Plural sentence + tool] tr^*(G [$\text{act}_{\text{transitive}}$]$_{\text{past}}$ an obj with a device, distributive): $\forall x\ (member(x, \text{G}) \rightarrow \exists E\ \exists y\ \exists z$ (core[S5.1.a](E, x+z, y) $\wedge$ $tool_{tp}$(z, E) $\wedge$ device(z))).

(g) [Plural sentence + tool] tr^*(G [$\text{act}_{\text{transitive}}$]$_{\text{past}}$ an obj with a device, collective): $\exists E\ \exists x\ \exists y$ (core[S5.1.a] (E, G+y, x) $\wedge$ $group_{tp}$(G+y, E) $\wedge$ $tool_{tp}$(y, E) $\wedge$ device(y)).

These schemata show that the following FO-sentence is included in every translation: $\exists E\ \exists x\ \exists y$ ($agent_{tp}(x, E) \wedge \text{act}_{\text{transitive}}(E) \wedge object_{tp}(y, E) \wedge \text{obj}(y) \wedge exist\text{-}time(E) < now$). This FO-sentence corresponds to tr(someone [$\text{act}_{\text{transitive}}$]$_{\text{past}}$ an obj) and means: For some E, x, and y, [x is an $agent_{tp}$ of E, E is an $\text{act}_{\text{transitive}}$, y is an $object_{tp}$ of E, y is an obj, and E is past]. In a simple singular sentence, the following condition is added to this basic FO-sentence: S is an atomic $agent_{tp}$ of E (see (S5.1.b)). Similarly, in a plural sentence in the collective reading, the following condition is added: G is a $group_{tp}$ for E (see (S5.1.e)). The distributive reading of a plural sentence "G [$\text{act}_{\text{transitive}}$]$_{\text{past}}$ an obj" interprets its content as follows: each member x of G performs an action expressed in simple sentence "x [$\text{act}_{\text{transitive}}$]$_{\text{past}}$ an obj" (see (S5.1.d)). In this distributive case, all individual actions are independently performed so that no collective agent exists in this context. An action using a device as a tool can be easily expressed by interpreting $agent_{tp}$ as an *extended* $agent_{tp}$ with the tool and adding conditions that u is a $tool_{tp}$ for E and that u is a device for a proper variable u (see (S5.1.c+f+g)). As you can see, all FO-translation schemata in (S5.1) are straightforward. The complexity of meaning of action sentences with extended and collective agents emerges from implications that some fundamental relations involve. Based on this consideration, the meaning of fundamental relations such as $agent_{tp}$, $tool_{tp}$, and $group_{tp}$ is axiomatically characterized in Th(AT).[8]

This paper provides a semantic analysis of action sentences based on Th(AT). However, we did not answer the following fundamental question: How do atomic agents realize a collective action? A task of this paper was to deliver a semantic basis for investigations on collective intentionality. The next step will be to develop a theory of collective intentionality based on Th(AT) and the presupposition that Th(AT) is shared by members of a linguistic community.[9]

Concluding Remarks

In this paper, I proposed a *four-dimensionalist axiomatic theory of actions and agents* (called Th(AT)) and analyzed collective actions and extended agents. In this paper, I have shown that Th(AT) is quite useful to investigate temporally extended complex (collective) actions. I have also suggested that tools extend the range of actions. In fact, inventions of tools have a potential to change societies and their environments. [10] [11]

Appendix 1

Here, I precisely describe some definitions and characterizations discussed in the main text.

8 The characterization of agency in Th(AT) gives only a necessary condition for agency. A description of intentionality is needed for a full characterization of agency.

9 A formal model of agents proposed in Nakayama (2022) might be useful to describe the relationship between agents and the society.

10 For discussions on the development of technology, see Nakayama (2016: Chapter 8).

11 I would like to thank two reviewers for many valuable suggestions.

(Ap1.1) General Extensional Mereology (GEM) (modified from Varzi (2019))
(a) [Reflexivity] $\forall x\ \mathrm{P}(x, x)$.
(b) [Anti-symmetry] $\forall x\ \forall y\ ((\mathrm{P}(x, y) \wedge \mathrm{P}(y, x)) \rightarrow x = y)$.
(c) [Transitivity] $\forall x\ \forall y\ \forall z\ ((\mathrm{P}(x, y) \wedge \mathrm{P}(y, z)) \rightarrow \mathrm{P}(x, z))$.
(d) [Overlap] $\forall x\ \forall y\ (\mathrm{O}(x, y) \leftrightarrow \exists z\ (\mathrm{P}(z, x) \wedge \mathrm{P}(z, y)))$.
(e) [Strong Supplementation] $\forall x\ \forall y\ (\neg\ \mathrm{P}(y, x) \rightarrow \exists z\ (\mathrm{P}(z, y) \wedge \neg\ \mathrm{O}(z, x)))$.
(f) [Unrestricted Fusion] For any formula $\varphi(x)$, $\exists x\ \varphi(x) \rightarrow \exists z\ \forall y\ (\mathrm{O}(y, z) \leftrightarrow \exists x\ (\varphi(x) \wedge \mathrm{O}(y, x)))$, when variables y and z do not occur free in $\varphi(x)$.
(g) [Notation for fusion] $\exists x\ \varphi(x) \rightarrow \forall z\ (z = \sigma x\ \varphi(x) \leftrightarrow \forall y\ (\mathrm{O}(y, z) \leftrightarrow \exists x\ (\varphi(x) \wedge \mathrm{O}(y, x))))$, when variables y and z do not occur free in $\varphi(x)$.
(h) [Proper Part] $\forall x\ \forall y\ (\mathrm{PP}(x, y) \leftrightarrow (\mathrm{P}(x, y) \wedge \neg\ \mathrm{P}(y, x)))$.
(i) [Sum] $\forall x\ \forall y\ (x{+}y = \sigma z\ (\mathrm{P}(z, x) \vee \mathrm{P}(z, y)))$.

(Ap1.2) Theory for four-dimensional event ontology (Th(4EO))
(a)In this paper, I express quantifications over different sorts of objects by relativizations. Let α and β be two sorts of variables and *Domain* be a *sub-domain predicate*. Then, $\forall\beta\ \varphi(\beta)$ and $\exists\beta\ \varphi(\beta)$ are used as abbreviations of $\forall\alpha\ (\mathit{Domain}(\alpha) \rightarrow \varphi(\alpha))$ and $\exists\alpha\ (\mathit{Domain}(\alpha) \wedge \varphi(\alpha))$, respectively. Here, I use *Space-time*, *Time*, *Event*, and *Music* as sub-domain predicates (see Table 7).

Sub-domain	Variables	Sub-domain predicates
Space-time objects	$S, S_1, S_2, \ldots$	*Space-time*
Time objects	$T, T_1, T_2, \ldots$	*Time*
Events	$E, E_1, E_2, \ldots$	*Event*
Musical compositions	$a, a_1, a_2, \ldots$	*Music*

Table 7. Variables and predicates for relativizations in Th(4EO)

(b) $\neg\exists\alpha\ (\mathit{Sub\text{-}domain}_1(\alpha) \wedge \mathit{Sub\text{-}domain}_2(\alpha))$ for any two different sub-domain predicates from {*Space-time*, *Time*, *Event*, *Music*}. Additionally, we require: $\forall\alpha\ (\mathit{Space\text{-}time}(\alpha) \vee \mathit{Time}(\alpha) \vee \mathit{Event}(\alpha) \vee \mathit{Music}(\alpha))$.
(c) GEM for *part*, GEM for part_T, and GEM for part_{st}.
(d) $\forall E\ \exists^{=1} S\ \mathit{occupy}_{st}(E, S)$.
(e) $\forall E\ \exists^{=1} T\ \mathit{occupy}_T(E, T)$.
(f) $\forall E\ \forall S\ (S = \mathit{st}(E) \leftrightarrow \mathit{occupy}_{st}(E, S))$.
(g) $\forall E\ \forall T\ (T = \mathit{exist\text{-}time}(E) \leftrightarrow \mathit{occupy}_T(E, T))$.
(h) $\forall E_1\ \forall E_2\ (\mathit{part}(E_1, E_2) \rightarrow \mathit{part}_{st}(\mathit{st}(E_1), \mathit{st}(E_2)))$.
(i) $\forall E_1\ \forall E_2\ (\mathit{part}(E_1, E_2) \rightarrow \mathit{part}_T(\mathit{exist\text{-}time}(E_1), \mathit{exist\text{-}time}(E_2)))$.
(j) [Relation *temp-part*] $\forall E_1\ \forall E_2\ (\mathit{temp\text{-}part}(E_1, E_2) \leftrightarrow\ \forall E_3\ (\mathit{exist\text{-}time}(E_3) = \mathit{exist\text{-}time}(E_1) \rightarrow (\mathit{part}(E_3, E_1) \leftrightarrow \mathit{part}(E_3, E_2))))$.
(k) [Function *temporal-part*] $\forall T\ \forall E_1\ (\mathit{part}_T(T, \mathit{exist\text{-}time}(E_1)) \rightarrow \forall E_2\ (\mathit{temporal\text{-}part}(E_1, T) = E_2 \leftrightarrow (\mathit{temp\text{-}part}(E_2, E_1) \wedge T = \mathit{exist\text{-}time}(E_2))))$.
(l) $\forall x\ \varphi(x)$ and $\exists x\ \varphi(x)$ are abbreviations of $\forall E\ (\mathit{Thing}(E) \rightarrow \varphi(E))$ and $\exists E\ (\mathit{Thing}(E) \wedge \varphi(E))$, respectively. Thus, *Thing* denotes a sub-domain of events.[12] I use $x, y, z, \ldots$ as variables for things.
(m) $\forall x\ \forall y\ \forall E\ (\mathit{part}_{tp}(x, y, E) \leftrightarrow \mathit{part}(\mathit{temporal\text{-}part}(x, \mathit{exist\text{-}time}(E)), \mathit{temporal\text{-}part}(y, \mathit{exist\text{-}time}(E))))$.
(n) [Spatial Part] $\forall E_1\ \forall E_2\ (\mathit{sp\text{-}part}(E_1, E_2) \leftrightarrow (\mathit{part}(E_1, E_2) \wedge \mathit{exist\text{-}time}(E_1) = \mathit{exist\text{-}time}(E_2)))$.
(o) $\forall T_1\ (\mathit{atomic}(T_1) \leftrightarrow \forall T_2\ (\mathit{part}_T(T_2, T_1) \rightarrow T_2 = T_1))$.
(p) $\forall t\ \varphi(t)$ and $\exists t\ \varphi(t)$ are used as abbreviations of $\forall T\ (\mathit{atomic}(T) \rightarrow \varphi(T))$ and $\exists T\ (\mathit{atomic}(T) \wedge \varphi(T))$, respectively.
(q) [Linearity] $\forall t\ \neg(t < t) \wedge \forall t_1\ \forall t_2\ \forall t_3\ ((t_1 < t_2 \wedge t_2 < t_3) \rightarrow t_1 < t_3) \wedge \forall t_1\ \forall t_2\ (t_1 < t_2 \vee t_1 = t_2 \vee t_2 < t_1)$.
(r) $\forall T_1\ \forall T_2\ (T_1 < T_2 \leftrightarrow \forall t_3\ \forall t_4\ ((\mathit{part}_T(t_3, T_1) \wedge \mathit{part}_T(t_4, T_2)) \rightarrow t_3 < t_4))$.
(s) $\mathit{Universe} = \sigma E\ \mathit{part}(E, E)$.
(t) *now* is an indexical that denotes the current time point.
(u) Th(4EO) consists of the requirements from (Ap1.2.a) to (Ap1.2.t).[13]

Th(AT) presupposes Th(4EO). In Th(AT), *agent*, *object*, *tool*, *member* are introduced as primitive relations.

(Ap1.3) Theory for actions and agents (Th(AT))).
(a) $\forall x\ \forall E\ (\mathit{agent}_{tp}(x, E) \leftrightarrow \mathit{agent}(\mathit{temporal\text{-}part}(x, \mathit{exist\text{-}time}(E)), E))$.
(b) $\forall x\ \forall E\ (\mathit{object}_{tp}(x, E) \leftrightarrow \mathit{object}(\mathit{temporal\text{-}part}(x, \mathit{exist\text{-}time}(E)), E))$.
(c) $\forall x\ \forall E\ (\mathit{tool}_{tp}(x, E) \leftrightarrow \mathit{tool}(\mathit{temporal\text{-}part}(x, \mathit{exist\text{-}time}(E)), E))$.
(d) $\forall E\ (\mathit{action}(E) \leftrightarrow \exists x\ \mathit{agent}_{tp}(x, E))$.
(e) $\forall x\ \forall y\ \forall E\ ((\mathit{agent}_{tp}(x, E) \wedge \mathit{agent}_{tp}(y, E)) \rightarrow x = y)$.
(f) $\forall x\ \forall E\ (\mathit{atomic\text{-}agent}_{tp}(x, E) \rightarrow \mathit{agent}_{tp}(x, E))$.
(g) $\forall x\ \forall y\ \forall E_1\ \forall E_2\ ((\mathit{atomic\text{-}agent}_{tp}(x, E_1) \wedge \mathit{sp\text{-}part}(E_2, E_1) \wedge \mathit{agent}_{tp}(y, E_2) \wedge \mathit{part}(y, x)) \rightarrow (y = x \wedge E_2 = E_1))$.
(h) $\forall x\ \forall E_1\ (\mathit{agent}_{tp}(x, E_1) \rightarrow \exists y\ \exists E_2\ (\mathit{part}(E_2, E_1) \wedge \mathit{part}_{tp}(y, x, E_2) \wedge \mathit{atomic\text{-}agent}_{tp}(y, E_2)))$.
(i) $\forall x\ \forall E\ (\mathit{extended\text{-}agent}_{tp}(x, E) \leftrightarrow (\mathit{agent}_{tp}(x, E) \wedge$

12 For characterizations of things, see Section 4 in Nakayama (2017). For example, a person is stipulated as the whole life of the person.
13 This description of Th(4EO) is based on Nakayama (2017). Nakayama (2017) proposes an extension of the axiomatic system in this paper.

$\neg$*atomic-agent*$_{tp}$(x, E))).

(j) $\forall x\ \forall E$ (*object*$_{tp}$(x, E) $\rightarrow \exists y$ *agent*$_{tp}$(y, E)) $\wedge\ \forall a\ \forall E$ (*object*(a, E) $\rightarrow \exists y$ *agent*$_{tp}$(y, E)).

(k) $\forall x\ \forall E$ (*tool*$_{tp}$(x, E) $\rightarrow \exists y$ (*agent*$_{tp}$(y, E) $\wedge$ *part*$_{tp}$(x, y, E))).

(l) $\forall x\ \forall y\ \forall E_1$ ((*agent*$_{tp}$(x+y, E_1) $\wedge$ *tool*$_{tp}$(y, E_1)) $\rightarrow$ $\exists E_2$ (use(E_2) $\wedge$ *sp-part*(E_2, E_1) $\wedge$ *agent*$_{tp}$(x, E_2) $\wedge$ *object*$_{tp}$(y, E_2))).

(m) $\forall x\ \forall E$ (*group*$_{tp}$(x, E) $\leftrightarrow$ *group*(*temporal-part*(x, *exist-time*(E)), E)).

(n) $\forall x\ \forall y\ \forall E$ ((*group*$_{tp}$(x, E) $\wedge$ *group*$_{tp}$(y, E)) $\rightarrow x = y$).

(o) $\forall x\ \forall E_1$ (*group*$_{tp}$(x, E_1) $\rightarrow \forall y$ (*member*$_{tp}$(y, x, E_1) $\leftrightarrow$ (*member*(*temporal-part*(y, *exist-time*(E_1)), *temporal-part*(x, *exist-time*(E_1))) $\wedge \exists E_2$ (*proper-part*(E_2, E_1) $\wedge$ *agent*$_{tp}$(y, E_2))))).

(p) $\forall x\ \forall E$ (*group*$_{tp}$(x, E) $\rightarrow \exists^{\geq 2} y$ *member*$_{tp}$(y, x, E)).

(q) $\forall E$ (*collective-action*$_{tp}$(E) $\leftrightarrow \exists x$ (*agent*$_{tp}$(x, E) $\wedge$ *group*$_{tp}$(x, E))).

(r) $\forall x\ \forall E_1$ ((*collective-action*$_{tp}$(E_1) $\wedge$ *group*$_{tp}$(x, E_1)) $\rightarrow E_1 = \sigma E_2\ \exists y$ (*member*$_{tp}$(y, x, E_1) $\wedge$ *part*(E_2, E_1) $\wedge$ *agent*$_{tp}$(y, E_2))).

(s) Th(AT) consists of the requirements in Th(4EO) and the requirements from (Ap1.3.a) to (Ap1.3.r).

Appendix 2

In this part, I sketch proofs of consistency of GEM, Th(4EO), and Th(AT).

(Ap2.1) [Proposition] In FO-Logic, the consistency of theory T can be proved by showing that there is a model for T.

PROOF. For FO-Logic, the strong completeness holds. This means: φ follows from T iff every model for T is a model for $\{\varphi\}$. Thus, $\varphi \wedge \neg\varphi$ does not follow from T iff there is a model for T that is not a model for $\{\varphi \wedge \neg\varphi\}$. However, according to the semantic definition of $\wedge$ and $\neg$, there is no model for $\{\varphi \wedge \neg\varphi\}$. Thus, T is consistent iff there is a model for T. Q.E.D.

(Ap2.2) [Proposition] Let S be a set of simple elements. We define $U(S) = P(S) - \{\varnothing\}$, where $P(S) = \{X: X \subseteq S\}$ and $\varnothing$ is the empty set. Then, structure $\langle U(S), I\rangle$ with $I(part) = \subseteq$ is a model for GEM, where I is an interpretation function.

PROOF. We can prove that $\langle U(S), I\rangle$ with $I(part) = \subseteq$ satisfies all axioms of GEM. Thus, this structure is a model for GEM. Q.E.D.

GEM is a subsystem of Boolean algebras. The following proposition suggests this fact.[14]

(Ap2.3) [Proposition] A model of Boolean algebras can be constructed from structure $\langle P(S), \subseteq\rangle$.

PROOF. Suppose that $\langle P(S), \subseteq\rangle$ is given. We introduce some functions through the following explicit definitions: $\forall a\ \forall b\ \forall c\ (c = a \cup b \leftrightarrow\ \forall d\ (d \subseteq c \leftrightarrow (d \subseteq a \vee d \subseteq b)))$, $\forall a\ \forall b\ \forall c\ (c = a \cap b \leftrightarrow \forall d\ (d \subseteq c \leftrightarrow (d \subseteq a \wedge d \subseteq b)))$, and $\forall a\ \forall b\ (b = \textit{complement}(a) \leftrightarrow \forall c\ (c \subseteq b \leftrightarrow \neg\ c \subseteq a))$. Then, we can easily prove that structure $\langle P(S), \cup, \cap, \textit{complement}, \varnothing, S\rangle$ satisfies all axioms of Boolean algebras. Q.E.D.

(Ap2.4) [Corollary] GEM is consistent.

PROOF. This proposition follows from (Ap2.1+2). Q.E.D.

(Ap2.5) [Proposition] Th(4EO) is consistent.

PROOF. At first, we construct a structure. Let S_T be the set of real numbers and $U(S_T) = P(S_T) - \{\varnothing\}$. We stipulate: For all T_1, T_2 with $T_1 \subseteq S_T$ and $T_2 \subseteq S_T$, [$T_1 <_T T_2$ iff [for all real numbers t_3 and t_4, if $\{t_3\} \subseteq T_1$ and $\{t_4\} \subseteq T_2$, then $t_3 < t_4$]]. A structure for time objects is defined as $\langle U(S_T), I_T\rangle$ with $I_T(part_T) = \subseteq$ and $I_T(<) = <_T$. For space-time objects, we define: $S_{st} = \{\ \langle s_1, s_2, s_3, t\ \rangle : s_1, s_2,$ and s_3 are real numbers and $t \in S_T\}$, $U(S_{st}) = P(S_{st}) - \{\varnothing\}$, and $I_{st}(part_{st}) = \subseteq$. Then, according to (Ap.2.2), $\langle U(S_T), I_T\rangle$ and $\langle U(S_{st}), I_{st}\rangle$ are models for GEM. For the sake of simplicity, we accept only one simple four-dimensionally extended trajectory $E_{\text{trajectory}}$. We stipulate: $U(E_{\text{trajectory}}) = \{E: E$ is a constituent of $E_{\text{trajectory}}\}$ and $I_E(part) = \textit{constituent-of}$. Then, we can prove that $\langle U(E_{\text{trajectory}}), I_E\rangle$ is a model for GEM by proving that this structure satisfies all axioms of GEM. Now, we introduce structure $\langle U(S_T) \cup U(S_{st}) \cup U(E_{\text{trajectory}}) \cup U(M), I\rangle$ with $I(\textit{Time}) = U(S_T)$, $I(\textit{Space-time}) = U(S_{st})$, $I(\textit{Event}) = U(E_{\text{trajectory}})$, $I(\textit{Music}) = U(M) =$ {Beethoven's spring sonata (BSS)}, $I(part_T) = I_T(part_T)$, $I(part_{st}) = I_{st}(part_{st})$, $I(<) = I_T(<)$, $I(part) = I_E(part)$, $I(occupy_{st}) = \{\ \langle E, s\rangle : s$ is the four-dimensional region occupied by E & $s \in U(S_{st})$ & $E \in U(E_{\text{trajectory}})\}$, and $I(occupy_T) = \{\ \langle E, T\rangle :$ T is the temporal region occupied by E & $T \in U(S_T)$ & $E \in U(E_{\text{trajectory}})\}$. You can prove that this structure satisfies all axioms of Th(4EO). Thus, according to (Ap2.1), Th(4EO) is consistent. Q.E.D.

A trajectory in a four-dimensional space can be interpreted as a thing. Such a thing represents a worm discussed in the worm theory, which is a version of the four-dimensionalism.

(Ap2.6) [Proposition] Th(AT) is consistent.

PROOF. In this proof, notions of Th(4EO) are used, where (Ap2.5) justifies this treatment. Now, let us consider a small world consisting of Mary, Tom, Mary's piano, and Tom's violin. Now, suppose that the only complete event in this world is their play of BSS and that all other events are parts of this complete event. Let TP(E_1, E_2) be an abbreviation of "E_1 is a temporal part of E_2". Now, let us define structure

14 Tsai (2009) gives a short overview of models of mereological theories and Boolean algebras.

$\langle U(S_T) \cup U(S_{st}) \cup U(E_{BSS}) \cup U(M), I_{BSS} \rangle$, where $U(M) = \{BSS\}$ and $U(E_{BSS})$ is the set of all events (including things) that deal with this play of BSS. I characterize the interpretation function I_{BSS} as follows.

(a) G_{BSS} = Mary+piano$_{Mary}$+Tom+violin$_{Tom}$, $E_{play\text{-}BSS} = E_{play\text{-}piano}+E_{play\text{-}violin}$, $E_{use\text{-}piano}$ is a part of $E_{play\text{-}piano}$, and $E_{use\text{-}violin}$ is a part of $E_{play\text{-}violin}$.

(b) $I_{BSS}(Thing)$ = {d: d is a mereological entity constructed from elements of {Mary, Tom, piano$_{Mary}$, violin$_{Tom}$}}, where the existence time of these entities is identical with the existence time of $E_{play\text{-}BSS}$.

(c) $I_{BSS}(atomic\text{-}agent_{tp}) = \{\langle Mary, E\rangle : TP(E, E_{use\text{-}piano})\} \cup \{\langle Tom, E\rangle : TP(E, E_{use\text{-}violin})\}$.

(d) $I_{BSS}(extended\text{-}agent_{tp}) = \{\langle Mary+piano_{Mary}, E\rangle : TP(E, E_{play\text{-}piano})\} \cup \{\langle Tom+violin_{Tom}, E\rangle : TP(E, E_{play\text{-}violin})\} \cup \{\langle G_{BSS}, E\rangle : TP(E, E_{play\text{-}BSS})\}$.

(e) $I_{BSS}(agent_{tp}) = I_{BSS}(atomic\text{-}agent_{tp}) \cup I_{BSS}(extended\text{-}agent_{tp})$.

(f) $I_{BSS}(object_{tp}) = \{\langle piano_{Mary}, E\rangle : TP(E, E_{use\text{-}piano})\} \cup \{\langle violin_{Tom}, E\rangle : TP(E, E_{use\text{-}violin})\}$ & $\{\langle BSS, E\rangle : TP(E, E_{play\text{-}BSS})\} \subseteq I_{BSS}(object)$.

(g) $I_{BSS}(tool_{tp}) = \{\langle piano_{Mary}, E\rangle : TP(E, E_{play\text{-}piano})\} \cup \{\langle violin_{Tom}, E\rangle : TP(E, E_{play\text{-}violin})\}$.

(h) $I_{BSS}(group_{tp}) = \{\langle G_{BSS}, E\rangle : TP(E, E_{play\text{-}BSS})\}$.

(i) $I_{BSS}(member_{tp}) = \{\langle Mary+piano_{Mary}, G_{BSS}, E\rangle : TP(E, E_{play\text{-}BSS})\} \cup \{\langle Tom+violin_{Tom}, G_{BSS}, E\rangle : TP(E, E_{play\text{-}BSS})\}$.

(j) $I_{BSS}(collective\text{-}action_{tp}) = \{E: TP(E, E_{play\text{-}BSS})\}$.

Then, you can prove that this structure satisfies all axioms of Th(AT). Thus, $\langle U(S_T) \cup U(S_{st}) \cup U(E_{BSS}) \cup U(M), I_{BSS}\rangle$ is a model for Th(AT). Then, according to (Ap2.1), Th(AT) is consistent. Q.E.D.

References

Bratman, M.E. (2014), *Shared Agency: A Planning Theory of Acting Together*, Oxford: Oxford University Press.

Davidson, D. (1967), 'The Logical Form of Action Sentences', in N. Rescher (ed.), *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh Press. Reprinted in Davidson (1980): 105-48.

Davidson, D. (1980), *Essays on Actions and Events*, Oxford, UK: Oxford University Press.

Durkheim, É. (1893), *De la Division du Travail Social*. Paris: P.U.F.

Gilbert, M. (2014), *Joint Commitment: How We Make the Social World*, Oxford, UK: Oxford University Press.

Hawley, K. (2020), 'Temporal Parts', E.N. Zalta (ed.) The Stanford Encyclopedia of Philosophy, URL = <https://plato.stanford.edu/archives/sum2020/entries/temporal-parts/>.

Linnebo, Ø. (2022), 'Plural Quantification', E.N. Zalta (ed.) The Stanford Encyclopedia of Philosophy, URL = <https://plato.stanford.edu/archives/spr2022/entries/plural-quant/>.

Ludwig, K. (2016), *From Individual to Plural Agency: Collective Action: Volume* 1, Oxford: Oxford University Press.

Ludwig, K. (2017), *From Plural to Institutional Agency: Collective Action, Volume* 2, Oxford: Oxford University Press.

Ludwig, K. (2018), 'Chapter 36: Actions and Events in Plural Discourse', in M. Jankovic and K. Ludwig (eds.) *The Routledge Handbook of Collective Intentionality* (Routledge Handbooks in Philosophy), London: Routledge.

Nakayama, Y. (2013), 'The Extended Mind and the Extended Agent', *Procedia Social and Behavioral Sciences*, 97: 503-10.

Nakayama, Y. (2016), *Beyond the Paradigm Theory: A Construction of an Evolutionary Theory for Science and Technology*, Tokyo: Keiso-shobo, in Japanese.

Nakayama, Y. (2017), 'Event Ontology based on Four-dimensionalism', *Department Bulletin Paper*, Graduate School of Human Sciences, Osaka University, 43: 175-92, <http://hdl.handle.net/11094/60581>.

Nakayama, Y. (2019), *From Philosophy of Language to Metaphysics: A new Development in Four-dimensionalism*, Tokyo: Keiso shobo, in Japanese.

Nakayama, Y. (2022), 'Games and Social Reality', IntechOpen, DOI:10.5772/intechopen.108066, in B. Sobota (ed.) (2023), *Game Theory – From Idea to Practice*, London: IntechOpen.

Searle, J.R. (2010), *Making the Social World: The Structure of Human Civilization*, Oxford, UK: Oxford University Press.

Sider, T. (2001), *Four-Dimensionalism: An Ontology of Persistence and Time*, Oxford, UK: Oxford University Press.

Tsai, H. (2009), 'Decidability of Mereological Theories', *Logic and Logical Philosophy*, 18: 45-63.

Tsai, H. (2013), 'Decidability of General Extensional Mereology', *Studia Logica*, 101: 619–36.

Tuomela, R. (2013), *Social Ontology: Collective Intentionality and Group Agents*, New York: Oxford University Press.

Varzi, A. (2019), 'Mereology', E.N. Zalta (ed.) The Stanford Encyclopedia of Philosophy, URL = <https://plato.stanford.edu/archives/spr2019/entries/mereology/>.

Notes to Contributors

1. All submitted papers are subject to anonymous peer-review, and will be evaluated on the basis of their originality, quality of scholarship and contribution to advancing the understanding of applied ethics and philosophy.

2. Papers should be at most 8,000 words, including references.

3. An abstract of 150-300 words and a list of up to 5 keywords should be included at the beginning of the paper.

4. Papers can be submitted any time of the year by e-mail to jaep@let.hokudai.ac.jp . If the authors wish their papers to be included in the next volume (to be published in March 2024), however, they are advised to submit their papers by September 15th 2023.

5. In-text references should be cited in standard author-date form: (Walzer 1977; Kutz 2004), including specific page numbers after a direct quotation, (Walzer 1977, 23-6).

6. A complete alphabetical list of references cited should be included at the end of the paper in the following style:

   Cohen, G.A. (1989), 'On the Currency of Egalitarian Justice', Ethics, 99 (4): 906-44.
   Kutz, C. (2004), 'Chapter 14: Responsibility', in J. Coleman and S. Shapiro (eds.), *Jurisprudence and Philosophy of Law*, Oxford, UK: Oxford University Press, 548-87.
   Walzer, M. (1977), *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, New York: Basic Book.

7. Accepted papers will be published in both web-based electronic and print formats..

8. The Editorial Board reserves the right to make a final decision on publication.