Asian Journal of Philosophy (2022) 1:51
https://doi.org/10.1007/s44204-022-00050-2

ORIGINAL ARTICLE

Logic of alternative-I

Yasuo Deguchi[1] · Takuro Onishi[1] · Ryota Akiyoshi[1,2,3] · Takashi Yagisawa[4] · Maiko Yamamori[5]

Received: 24 May 2022 / Accepted: 7 October 2022 / Published online: 15 November 2022

Abstract
This paper aims to construct a logic of *alternative-I* that provides a proper conceptual framework for talk of possible-I in decision-making context, and thereby solves what we call *the paradox of possible-I*. The model of our logic, Alt-I model, is an adaptation of N. Belnap's *branching-time* model, and the STIT (*see to it that*) operator defined on the model serves to represent choices and decisions made by actual and counterfactual agents. We conclude this paper by discussing the application of Alt-I model to the case of digital *twins*, digital copies of a person.

Keywords Possible-I · Alternative-I · Possiblia · Branching-Time · STIT · Digital twin

1 Introduction

This paper aims to construct a logic of *alternative-I* that provides a proper conceptual framework for talk of *possible-I* in decision-making context, and thereby solves what we call *the paradox of possible-I*.

Each of us sometimes think of (imagine, or fantasize) various possible-I's, i.e., entities I myself could be, which are different from the entity I actually am. But the relation between actual-I and possible-I may incubate an ambiguity or paradox that,

Dr. Maiko Yamamori passed away on April 7, 2022.

✉ Yasuo Deguchi
ydeguchi@bun.kyoto-u.ac.jp

Takuro Onishi
takuro.onishi@gmail.com

1 Graduate School of Letters, Kyoto University, Kyoto, Japan

2 Global Research Institute, Keio University, Tokyo, Japan

3 Institute for Advanced Study, Waseda University, Tokyo, Japan

4 California State University, Northridge, CA, USA

5 Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan

while they are distinct in some respects, the possible-I should be myself, so be identical to myself. In short, the actual-I and a possible-I are distinct but identical.

This paradox is a special case of a more general problem of the relation between an actual object and the corresponding possibilia (possible objects). For example, when, about the omelet I cooked this morning, I regretfully ponder that it would have been better if I hadn't made a mistake with the salt, I think of a possible omelet that is distinct from the actual one, but still exhibits a possible way in which the omelet could have been. So they should be identical, but at the same time not.

Several attempts have been made to dispel or parameterize away this apparent paradox. In those attempts, the relation between the actual and possible-I's is interpreted either as the classical identity in the Kripkean theory of transworld identity, as nonidentity relation, e.g., counterpart relation in the Lewisian theory or a relation between modal parts of a mereological whole in Yagisawa's five-dimensionalism. We will present an alternative notion of identity and distinctness by focusing our attention on possible-I in decision-making context rather than possibilia in general.

There are typical situations in which each of us will come to think of possible-I. Among them is that of decision-making, where we face some alternatives for the future course of our lives, and choose only one of them, leaving the others as merely counterfactual. Before or after making such a decision, each of us will be inclined to think of possible-I or what I would be if I chose one option rather than the others. Let us call possible-I in such a decision-making context *alternative-I.*

With this notion of alternative-I, the answer to the paradox of possible-I is straightforward. It is plain that an alternative-I is distinct from the actual myself since the former is a counterfactual life that I did not choose. Nonetheless, one will take it to be *my* alternative-I because it is a life that branched off from my life at some moment of decision. In other words, it shares with me an initial part of life. In this sense, an alternative-I is distinct from but identical with me.

In what follows, we substantiate the notion of alternative-I by providing with a formal model that we call the *Alt-I model*. As expected, alternative-I's of a person form the structure of *tree*, which represents branching-off of their courses of life. Then, one observes that there is a *degree of identity*. That is, we would say that an alternative-I x is *more identical* with the actual one a *than* another alternative-I y if x branches off from a at a (temporarily) later moment than the one where y does. Thus, the tree structure yields a topology among the actual- and alternative-I's.

Our Alt-I model is a branching *self* model, so to speak. It is an adaptation of Nuel Belnap's *branching-time (BT) model* that represents the branching structure of time in which many possible futures branch from a determinate past. The semantics of STIT (*see to it that*) operator defined on the BT-model is also transplanted to the Alt-I model and enables us to make sense of the context of decision-making where the notion of alternative-I comes into play.

Finally, as an application of the Alt-I model, we consider *digital twin as alternative-I*. Branching of self has long been merely a matter of sci-fi thought experiments. But now we are witnessing the emergence of technological possibilities or even feasibilities to realize it. Among those possibilities is the technology of digital twin: digital duplication of an actual thing, including an actual person. A digital twin of a person is a digital copy or simulation of the person that may autonomously engage in various aspects of our daily life as our proxy. If this should happen, and it is a real possibility that it will, in the near future, it will inevitably bring about many social issues, including ethical and legal challenges. On the other, a digital twin of a person has no peculiarity that cannot be found in other examples of alternative-I such as a parallel world character in fiction. So, taking into account its potential to raise real issues, we take digital twins as examples of alternative-I who inhabit the actual world and show how they are represented in the Alt-I model, without any loss of generality.

This paper will proceed as follows. In Section 2, we first have a brief overview of some other approaches to the paradox, i.e., the Kripkean, the Lewisian, and the Yagisawan theories. Then, the notion of alternative-I, possible-I in the context of decision-making, is presented, and compared with the above three approaches, and Kurt Lewin's genidentity, which has a similar structure to alternative-I. In Section 3, we introduce our logic of the Alt-I model by modifying Belnap's BT model, and discuss its application to the case of digital twins.

2 Alternative-I

2.1 Three approaches to possibilia

Here we will consider the Kripkean, the Lewisian, and the Yagisawan theories as the prevailing approach to the paradox of possible-I or the problem of possiblia in general.

In Kripke's stipulative constructivist conception of possible worlds or situations, a possible world is constructed by our stipulation as the world in which an actual individual, say Richard Nixon, has some accidental properties he actually lacks, say losing the 1968 US presidential election, while retaining all his essential properties (Kripke, 1980). In this view, the actual individual and its counterfactual counterpart are trivially identical because the very construction of the nonactual possible world in which the latter exists presupposes the identity relation to the former.

The theory can easily solve the paradox by introducing the distinction between essentials and accidentals, in such a way that the numeral identity of the actual winner Nixon and the possible loser Nixon is secured by sheer stipulation as long as they differ only in accidental properties.

On the other hand, according to the Lewisian modal realism, possible worlds are mereological wholes consisting of the individuals that bear some natural relation (typically, spatiotemporal relatedness) to one another (Lewis, 1968). So every individual exists only in one world; no individual in one world is transworldly identical with any individual in another world. Therefore, the relation between the actual-I and a possible-I is not transworld identity any longer. Instead, the relation is now interpreted as *counterpart relation*, i.e., the greatest similarity relative to a given possible world: a possible-I is the most similar individual, among all sufficiently similar entities in its possible world, to the actual-I.

Now, the paradox is singlehandedly solved by replacing one of its horns, i.e., identity, with the greatest similarity: the acutal-I and a possible-I are not identical but similar with each other in the greatest degree relative to a possible world.

The counterpart relation does not satisfy the uniqueness condition, e.g., the actual-I can be equally similar in the greatest degree to two or more individuals in a given possible world. So the relation can fork (from an actual-I to two or more possible-I's) or fuse (from the two or more possible-I's to the actual-I).

According to the Yagisawan five-dimensionalism, all possible individuals (including actual individuals) are spread in logical space, i.e., have modal parts in one or more possible worlds (Yagisawa, 2017). The spatiotemporally extended four-dimensional I in the actual world (actual-I) is a part of the modally extended mereological whole, a five-dimensional individual (the whole I) spread out in logical space at large. A possible-I is also another part of the same five-dimensional individual in another possible world. Now, the relation between the actual-I and a possible-I is interpreted as transworld sharing of a common mereological whole, or the relation between different modal four-dimensional parts that share the same five-dmensional whole.

Five-dimensionalism can also parametrize away the paradox by introducing a parameter of the whole and its parts. The actual-I and a possible-I are identical in that their whole is identical, while being not identical in that they are different parts of the whole.

Those theories are about possiblia in general, and not intend to tackle the paradox of possible-I specifically. If we focus on possible-I, and think about the situations in which possible-I's are the subject of serious consideration, then a different approach suggests itself.

2.2 Alternative-I: possible-I in the context of decision-making

It is philosophically interesting that we can imagine a very wide range of possibilities or ways we could be. But under what circumstances does it really matter to our lives to think about possibilities at all? The moments of decision-making will definitely be included in such circumstances. From what to eat for lunch today to what profession to choose, one shapes one's life by making a variety of decisions. So it is sensible enough to focus our attention on possible-I considered in decision-making situations, which we call *alternative-I*.

Each time we make a decision, we examine each of the options available to us at the moment and imagine how our lives will progress when we choose it. The chosen one is the actual-I's life. The others are left as counterfactual possibilities. One also recollects past decisions one made and rethink what one's life would have been like if an alternative path were chosen. One's actual life, the actual-I, is surrounded by the alternative-I's that were, are, or will be possible but not realized in this actual world.

This picture suggests that the structure formed by actual- and alternative-I's should be (represented as) a *tree* whose root is the moment of birth of the actual-I and among whose branches the distinguished one is the actual-I and the others alternative-I's. At a moment of decision, we are like standing at a fork and choosing one of the paths that extend from it. The path I myself choose is my actual life, the actual-I, and an alternative-I chooses and goes along a different path. *Fission* or *branching* towards the future is the basic structure of alternative-I[1].

It is to be noted here that, as can be seen from the tree structure, each alternative-I shares with the actual-I a history up to the moment of their branching off. There would be no substantial point in considering a possible future to choose if it were not *my* possible future, i.e., a possible continuation of the actual-I's course of life up to the moment.

It should now be clear how the paradox of possible-I should be solved in the context of decision-making. Alternative-I's are obviously distinct from the actual-I, since they are counterfactual. They are different courses of life from the one I actually take. Nonetheless, they can be taken to be *my* alternative lives since each of them is conceived as a possible continuation of some initial part of my life. This explains the identity and distinctness of actual- and alternative-I's. They are identical in that they share an initial part of their histories, and distinct in that they branch off from each other at some point.

2.3 Comparison with previous theories

We are not arguing that the theory of alternative-I is better than the previous theories concerning possibilia. It is only intended as a proposal that will fit the specific context of decision-making. But comparing it with the previous theories will make clearer the significance of the notion of alternative-I.

It may well be said that an alternative-I is transworldly identical with the actual-I as in the Kripkean theory except for that they need not share the essence. What makes them (transworldly) identical is the history they share. For example, suppose that, whereas you actually entered University *A* after graduating from high school, you might have made a decision to enter University *B* instead. The alternative-I who entered University *B* is *your* alternative because it shares the history up to the high school graduation, which should not be your essence.

We are also open to the view of alternative-I as a variant of Lewisian counterpart. A difference is that alternative-I's need not satisfy the greatest similarity condition, much less the uniqueness. Rather, we would like to emphasize that the degree of similarity or identity among actual- and alternative-I's is cashed out in the tree structure. Suppose that, last year at the age of forty, you had a choice between a job at University *A* and a job at University *B*, and chose *B* this time. Now, compare the alternative-I working at University *A* now and the one who entered University *B* about 20 years ago and spent their life up to the age of forty. You would feel the former is closer to yourself and the latter perhaps almost an utter stranger. In general, an alternative-I that diverges from the actual-I at a later point than another alternative-I does may be said to be *more identical with* the actual-I than the one

[1] This implies that we are excluding *fusion* of alternative-I's while allowing fission. We take this to be natural in the context of decision-making, although fusion might be possible in some different context.

that diverged earlier[2]. Thus, the tree structure of actual- and alternative-I's induces a *topology* among them with respect to the degree of identity.

Also, the theory of alternative-I can be viewed as a specification of the Yagisawan five-dimensionalism. The tree is a mathematical representation of the mereological whole (the whole-I) consisting of actual- and alternative-I's. It is a specification in that we limit our consideration about possibility to the context of decision-making. This specification brings in more structure to the mereological whole, that is, the branching structure and the topology with respect to identity.

Finally, we briefly mention the notion of genidentity as one of the approaches to branching (and fusing) identity. Kurt Lewin proposed a logical formulation of generative relation, in which an entity is generated from another, in terms of a non-classical, diachronic identity, *genidentity*, in the way that a and b are genidentical *iff* a is generated from b (Lewin, 1922, 1923; Padovani, 2013).

Genidentity can branch and fuse as well. Take a branching case. Suppose that two descendants are generated from a common ancestor. While each of the two descendants is genidentical to the ancestor, the descendants are not genidentical with each other. So the genidentity relation branches from the ancestor to the two descendants. On the other hand, fertilization of an unfertilized ovum by a sperm is an example of fusion. While the unfertilized ovum, as well as the sperm, is taken to be genidentical to a fertilized ovum, the unfertilized ovum and the spam are not genidentical. Here two instances of the genidentity relation, the one between the unfertilized ovum and the fertilized ovum, and the other between the sperm and the same fertilized ovum, converge at the fertilized ovum.

Genidentity (at least its original version) is not the relation between actual and possible entities. Thus, it does not purport to solve the paradox of possible-I. But an extension or modification to suit the theory of alternative-I will be straightforward. Notice that genidentity is the relation between moments, which are states of actual- and alternative-I's *at some temporal instants*, which means that it is a different type of relation from the identity relation between actual- and alternative-I's which we consider as temporally extended wholes. This is not an essential difference since they can be translated to each other in an obvious way (for branching cases).

How about fusion allowed for genidentity? We do not consider fusion among alternative-I's for the moment because, even if it is metaphysically possible, it would not matter to our decision-making. When you start conceiving a life with digital twins and conceptualizing them as alternative-I's inhabiting the actual world as we will see later in this paper, a fusion of digital twins or an "absorption" of twins to the actual-I would perhaps be considered[3].

[2] The degree of identity need not be estimated with respect to the actual-I. That is, we consider, rather than the two-place relation "b is more identical with the actual-I than c is," the three-place relation "b is more identical with a than c is" defined for any triple $\langle a, b, c\rangle$ of alternative-I's, one of which may be actual.

[3] What is fusion? On what condition can two agents be said to be fused with each other? The digital twin technology can choose one or another condition or definition of fusion. For instance, two agents can be regarded to be fused with each other only when they come to share all their memories. Or the technology can set a condition for the case of fusion between a real agent and a digital one: the fused agent should be real rather than digital, and inherit only the physical condition of the former rather than the latter.

3 Formal model of alternative-I

Now that we have given an outline of the notion of alternative-I, this section presents a formal model of alternative-I (Alt-I model). It is an adaptation of Belnap's branching-time model (BT-model), and logic of action and agency developed on the basis of the BT-model. In what follows in this section, we first introduce Belnap's framework following Belnap et al. (2001, Ch.2) and present how to reinterpret it as a model of alternative-I[4].

3.1 Branching time

The basic structure of BT-model is a tree $\langle T, < \rangle$[5], where T is a non-empty set and $<$ is a binary relation on T that is irreflexive, transitive, and non-backwards-branching. The elements of T are called *moments*, possible states of (a part of) a world at some time instants, and $<$ is a temporal ordering between moments. The non-backwards-branching (branching-into-future) property of $<$ represents the determinacy of past and indeterminacy of future.

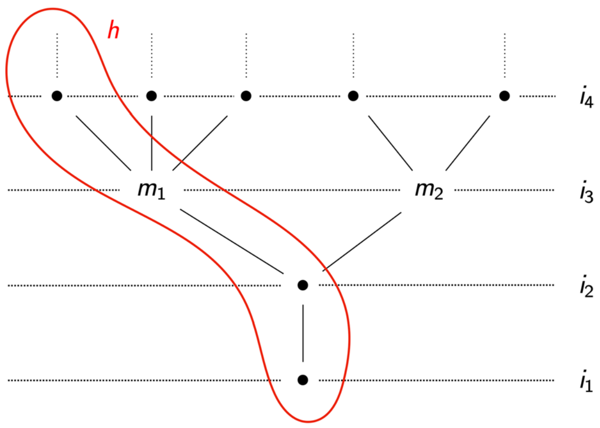
A maximal linearly ordered subset of T is called a *history*, representing a (possible) complete temporal development of a world. We denote the set of histories over T by $\mathcal{H}$, and for each moment $m \in T$, the set of histories passing through m by H_m. Thus, $H_m = \{h \in H \mid m \in h\}$.

We may add to BT-model one more component called *instants*, for which we write $\mathcal{I}$. It is a partition of the moments in the model into "contemporaneous" ones belonging to different histories. So it is assumed (1) that to an instant belongs at most one moment from a history, and (2) that instants respect the temporal ordering $<$ between moments, that is, if $m_0 < m_1$ in $h \in H$, $m'_0, m'_1 \in h' \in H$, $m_0, m'_0 \in i_0 \in \mathcal{I}$, and $m_1, m'_1 \in i_1 \in \mathcal{I}$, then $m'_0 < m'_1$. We define i_m to denote the instant that moment m belongs to. Hence, $m' \in i_m$ means m' is contemporaneous with m.

Example 1 The basic structure of a BT-model with instants can be described as below. Moments, nodes of the tree structure, are arranged along the timeline of instants (i_1, i_2, i_3, i_4). Moment m_1 belongs to history h and hence h belongs to H_{m_1}, but $h \notin H_{m_2}$. m_1 and m_2 are contemporaneous since both are at instant i_3 ($i_{m_1} = i_{m_2} = i_3$).

[4] It would be interesting to see how logics of action and agency like Belnap's and ours relate to decision theory. Since this is not the place to discuss it, we would like to just refer to Bartha (2014) as one of the works dealing with it. In the paper, Bartha presents a system of deontic logic based on Belnap's logic as a kind of *proto-decision theory*, which offers rigorous analysis of foundational notions in decision theory such as causation and choice in an illuminative way.

[5] In Belnap's original formulation, $\langle T, < \rangle$ may be a tree-like structure, a non-backwards-branching poset possibly without a root.



Now, let us call $\langle T, <, H, \mathcal{I} \rangle$ a *BT-frame*, and $\langle T, <, H, \mathcal{I}, \Vdash \rangle$ a *BT-model* on a BT-frame $\langle T, <, H, \mathcal{I} \rangle$ where $\Vdash$ a *forcing relation* or *valuation* over the pairs of a moment and a propositional variable:

$$m \Vdash p$$

means that p holds at moment m[6]. Complex formulas formed with classical connectives $\wedge, \vee, \supset, \neg$ are evaluated in the classical way.

3.2 Action, agency, and *STIT*

With a foundation of BT-model, Belnap develops a logic of action and agency giving a formal semantics for the linguistic form called *STIT*, i.e., "an agent α *sees to it that A*."[7] It may be denoted by a kind of necessity operator $\Box_\alpha$ indexed by agent α. Belnap interprets a STIT formula $\Box_\alpha A$ to mean

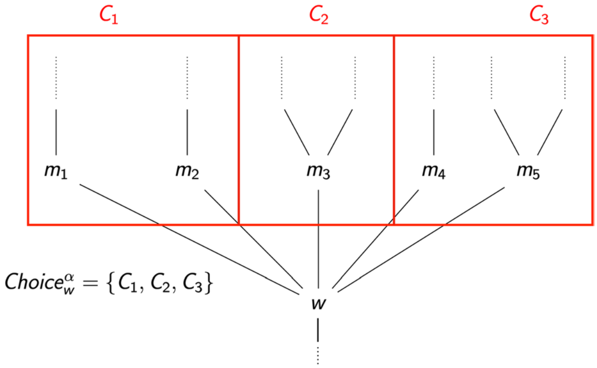
> the present momentary fact A is guaranteed by a prior choice of the agent α. (Belnap et al., 2001, 2A.2)

Now, let us have a BT-model equipped with a set of *agents*, say $\mathcal{A}$, and *Choice*, which is a function of agent and moment. Given agent α and moment m, $Choice^\alpha_m$, which we call the *choice set for agent α at moment m*, is a partition of the set H_m of

[6] In Belnap's original definition, formulas are evaluated relative to a *pair of moment and history* denoted by expression like m/h, where $h \in H_m$. Thus, $m/h \Vdash p$ says that a propositional variable p is true at moment m considered as a part of history h. This pair-form valuation serves to give truth definitions for some future tense operators. Since we are not concerned with tense in this paper, we take a valuation to simply be a function from pairs of a propositional variable and a moment like ordinary modal logics.

[7] In the literature, two kinds of STIT are considered: achievement STIT and deliberative STIT. Here we are dealing only the former.

histories developing through m[8]. It represents a set of possible choices or actions available to α at moment m. Agent α is supposed to choose one option from the choice set and thereby abandon the other options. Note that the option chosen is an equivalence *class* in the choice set. The future for the agent is still not determined uniquely. The diagram below shows an example of choice set, a partition of histories developing from moment w.



We say that moments m_0 and m_1 are (α/m) *choice-equivalent*, equivalent with respect to α's choice at moment m, and write $m_0 \sim^{\alpha}_{m} m_1$ if they belong to the same instant and the same equivalent class in choice set $Choice^{\alpha}_{m}$.

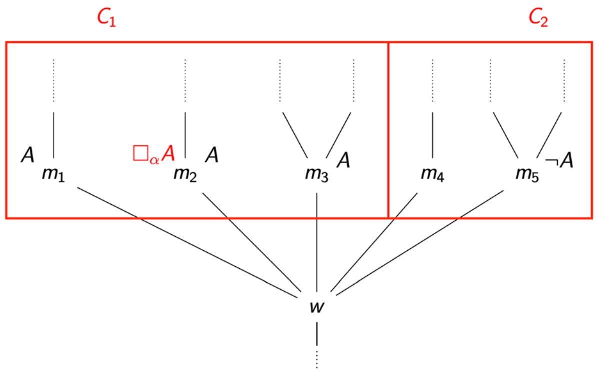
Let us explain how STIT is interpreted with this machinery. On a BT-model, the semantic clause for STIT is given as follows.

$$m \Vdash \Box_{\alpha} A \iff \exists m_0 < m \text{ such that}$$
$$(1)\ \forall m_1 : m_1 \sim^{\alpha}_{m_0} m \Rightarrow m_1 \Vdash A$$
$$(2)\ \exists m_2 \in i_m : m_2 \nVdash A.$$

When $m \Vdash \Box_{\alpha} A$, there is a moment m_0 of decision temporally prior to m, where agent α made a decision to guarantee that A holds at m: According to the first clause, A holds at every moment that are choice-equivalent with m (including m itself), which means the choice they belong to is the one which makes A true. The second clause is to make sure that the choice is not trivial in the sense that A holds every moment contemporaneous with m and hence there is no point in *choosing* A.

Example 2 In the following model, we assume that $Choice^{\alpha}_{w} = \{C_1, C_2\}$ as indicated. Then, $m_2 \Vdash \Box_{\alpha} A$. For (1) A holds at all moments that is choice-equivalent to m_2 (those in the same box with m_2), and (2) there is a moment, m_5 where A fails to be true.

[8] If $h, h' \in H_m$ are not divided at m, i.e., $\exists m' > m : m' \in h \cap h'$, then they should belong to the same equivalent class.



This formal explanation of STIT can be applied and extended in many ways. Among them, we are particularly interested in the extension to *joint STIT*, which formalizes joint action or decision by a group of agents (Belnap et al., 2001, Ch.10). There are at least two ways to understand joint action or decision in general: one is "distributive" action and the other is "collective" one. The former would be analyzed as conjunction of single agent STIT's like $\Box_\alpha A \wedge \Box_\beta A \wedge \Box_\gamma A$. For joint action by a group of agents taken collectively, joint STIT operator is introduced.

The notion of choice-equivalence is extended to cover multiple agents. Let $X \subseteq \mathcal{A}$ be a non-empty set of agents. Then for any $m, m_1, m_2 \in T$, we define $m_1 \sim_m^X m_2$ (equivalent with respect to the collective choice by X at m) if $m_1 \sim_m^\alpha m_2$ for each $\alpha \in X$. Thus, "we define a possible choice for [X] as a whole to be the intersection or "combination" of all the individual possible choices"(Belnap et al., 2001, Ch.10c.2).

Then, the semantic clause for joint STIT operator $\Box_X$ (a group X of agents sees to it that...) is defined as follows.

$$m \Vdash \Box_X A \iff \exists m_0 < m \text{ such that}$$
$$(1)\ \forall m_1 : m_1 \sim_{m_0}^X m \Rightarrow m_1 \Vdash A$$
$$(2)\ \exists m_2 \in i_m : m_2 \nVdash A.$$

3.3 Reinterpreting BT-model as model of alternative-I

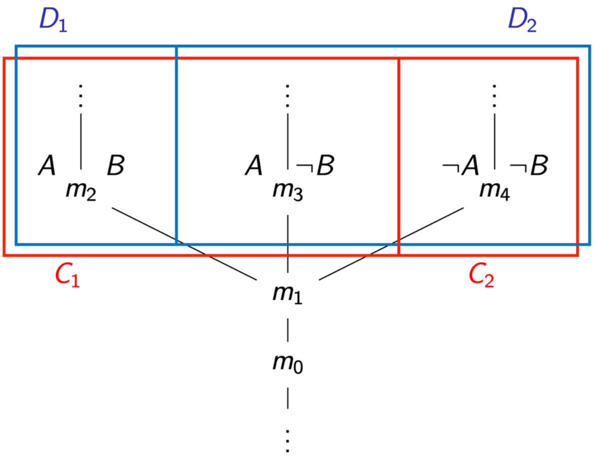
Let us now turn to our proposal. As remarked earlier, we understand a person's alternative-I's as forming a tree. We take a BT-model $\langle T, <, H, \mathcal{I}, \Vdash \rangle$ as representing a set of possible developments of a person's life, not of the world. Each history, $h \in H$, is an alternative-I[9], and a moment $m \in h \in H$ is a state of an alternative-I h at some instance. The forcing relation $m \Vdash A$ is now taken to state that an alternative-I at moment m has property A. We call the structure $\langle T, <, H, \mathcal{I}, \Vdash \rangle$ an *Alt-I model* under this interpretation.

9 Hereafter we use "alternative-I" as including the actual-I. Indeed, from the perspective of a (counterfactual) alternative-I, the actual-I is their (counterfactual) alternative-I.

When it comes to the issue of decision-making, it is moment, rather than alternative-I, that plays a remarkable role in the Alt-I model. An alternative-I is the end product, so to speak, of choices and actions that are made, and they are made at some point in its life when it is still in progress. So the *agent* of choices and actions should not be an alternative-I as a whole, completed life of a person, but an alternative-I at *some moment*. Then, in any Alt-I model, we may simply set $\mathcal{A} = T$, i.e., agents are moments[10].

Now, in Alt-model, a choice set is of the form $Choice^{m_0}_{m_1}$ where $m_0, m_1 \in T$, representing the set of options available at moment m_1, from which agent m_0 is going to choose. This makes sense if $m_0 \leq m_1$, in which case we suppose a situation like a person at moment m_0, as an agent of choice, considering what to do at a future moment m_1 or just now (when $m_0 = m_1$).

Let us see an example. In the figure below, we assume $Choice^{m_0}_{m_1} = \{C_1, C_2\}$ and $Choice^{m_1}_{m_1} = \{D_1, D_2\}$[11]. Suppose that you as a high school student (at m_0) are considering whether to enter graduate school (A) or to start a career in industry ($\neg A$) after graduating from a university (at m_1). Also, suppose that you are now at m_3. Now, it holds that $m_3 \Vdash \Box_{m_0} A$. The present fact that you are a graduate student is guaranteed by a prior choice at m_0, a choice by you as a high school student.



On the other hand, neither $m_3 \Vdash \Box_{m_0} B$ nor $m_3 \Vdash \Box_{m_0} \neg B$. Let us read B as "living in city b." You, as a high school student, did not make any decision whether to

[10] We may also take as an agent an initial segment of a branch in T, i.e., $\{x \mid x < m \text{ or } x = m\}$ for some $m \in T$. This represents a life history of an alternative-I up to the moment m, and, in this interpretation, a choice is made by the alternative-I that has lived its life up to m. Mathematically, a moment m determines an initial segment of an branch uniquely and vice versa, so we may define the set of agents in either way.

[11] Belnap imposes, on *Choice* function, the *independence constraint* to the effect that "no choice by [agent] α_1 can make it impossible for [agent] α_2 to make a *simultaneous* choice" (Belnap et al., 2001, 2A.2) (emphasis added). That is, for any moment m, agents α, β, and options $C \in Choice^{\alpha}_{m}$, and $D \in Choice^{\beta}_{m}$, $C \cap D \neq \emptyset$. As you can see from the present example, we do not require the independence. In an Alt-I model, $Choice^{m_0}_{m_1}$ represents the choice made at moment (= by agent) m_0 about what to do at m_1. The choice may not be simultaneous with the choice by other agents; hence, the independence should not be required. For the case of digital twin we look at later, even simultaneous choices may not be independent due to the power relation between the actual-I and their (substitutional) digital twins.

live in b or not. Now, $m_3 \Vdash \Box_{m_1} \neg B$. It is when you graduated at m_1 that you decided to live outside b[12].

With this example, we saw decisions by a person at different moments. Moreover, by using joint STIT, one can also express decisions by a continuously existing person. Indeed, it holds $m_3 \Vdash \Box_{m_0,m_1}(A \wedge \neg B)$. To enter a graduate school outside b $(A \wedge \neg B)$ is a joint decision by you at m_0 and at m_1.

Though $Choice^{m_0}_{m_1}$ works well in the above example, a choice set hardly makes sense if $m_0 \nleq m_1$, i.e., if $m_1 < m_0$ or they are not comparable. For then we would consider, with $Choice^{m_0}_{m_1}$, decisions on past actions or decisions for (counterfactual) alternative-I's actions. In view of this, we require that, in any Alt-I model, *Choice* satisfy the following constraint:

$$\text{if } m_0 \nleq m_1, \text{ then } Choice^{m_0}_{m_1} = \{H_{m_1}\}.$$

Note that $Choice^{m_0}_{m_1}$ is, in general, a partition of H_{m_1}. Choosing an option at m_1 is choosing a class in the partition and discarding the other classes (options). Then, that $Choice^{m_0}_{m_1} = \{H_{m_1}\}$ means that no option is discarded, and hence the choice by m_0 at m_1 is not substantial. Indeed, if $Choice^{m_0}_{m_1} = \{H_{m_1}\}$, then moment m_1 never serves as a moment of decision for $\Box_{m_0}$[13].

3.4 Digital twins

In this section, as an application of the Alt-I model, we discuss how our decision-making would be if we became to live with our digital twins. By a digital twin, we mean an autonomous intelligent agent created as a digital copy of a person. It is supposed to have transplanted into it (at least a large part of) the person's biological information, memories, tastes, thinking styles, and so on. It may be implemented as a mere digital avatar on a computer screen, or as a kind of robot with a physical body. Either way, it will be able to behave as an autonomous agent of choice and action. Among various artificial agents such as robots, AI assistants, and androids,

[12] The form $m \Vdash \Box_{m'} A$ ascribes to moment (i.e., state of an alternative-I) m the property that the present fact A about it is due to the decision by agent m'. In other words, the STIT form is *retrospective* in the sense that it looks back on past decisions in terms of present facts. It will be possible to consider a *prospective* form, which we may write like $m \Vdash \odot_{m'} A$ or $m \Vdash \odot_i A$, which ascribes to m the property of deciding to do A at $m' \geq m$ or instant i in the future (here m is an agent rather than moment, and m' a moment). Giving semantics to this form is our future task. We thank an anonymous reviewer for raising this point.

[13] The proposed Alt-I model might appear to be incompatible with counterpart theory, but the appearance is illusory. In our presentation, we deliberately conflate branching and divergence in Lewis's senses. Branching involves one common initial segment, whereas divergence involves two duplicate initial segments. We have chosen the language of branching for simplicity's sake, but it can be systematically replaced by the language of divergence, which Lewis prefers. See Lewis (1986, 206-207, also 70-71 and 125).

digital twins are peculiar in that they are not only a new kind of agents, but also are *our* copies. In a sense, they are *ourselves*.

As you notice, we may take a digital twin of a person as their alternative-I. Since it is created by copying the history of the person up to the moment, it is identical with its "organic" original according to the identity criterion we proposed in the previous section. Of course, it is also distinct from the organic twin since it would live its own life after the moment. Thus, their lives are alternative to each other, sharing the initial history and diverging at some point.

Yet, digital twins have a different modal status from alternative-I's we have considered so far in this paper. My digital twin is an alternative-I that is living a counterfactual life different from my actual life, so it is like a possibile or counterpart of mine in a possible world. On the other hand, it is also supposed to interact with myself and other people in the actual world. Thus, digital twins are, as it were, counterfactual possible-I's living in the actual world. In what follows, leaving the detailed characterization of the modal status of digital twin for another occasion, we only consider what changes are brought in to the Alt-I model when digital twins come to existence.

Digital twins will play various roles in a society. One can conceive of, for instance, a *dialogical* digital twin, which lives with the original person as the closest dialogue partner with mutual understanding with the original person better than any friend or family member. Also, there will be a *substitutional* digital twin, that is, one designed to do some jobs, like writing e-mails or attending business meetings, on behalf of its original person as they intend, and so behave like a subordinate rather than a partner or friend of the original person. We propose some constraints on *Choice* in the Alt-I model that will characterize the relationship between a person and their substitutional digital twin[14], and the difference between digital twins and counterfactual alternative-I's living in possible worlds.

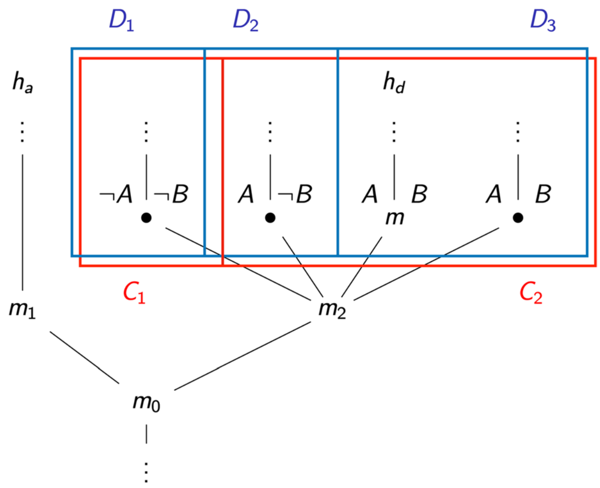
For simplicity, we consider a case where, in a given Alt-I model, there is just one (substitutional) digital twin $h_d \in H$ created at moment m_0 by copying the actual and original person $h_a \in H$. The following discussion will be extended easily to more complex cases where a person has multiple digital twins or even a digital twin has its twins.

First, since digital twins are supposed to interact with us in the actual world, and the original person can make decisions for them, we propose to loosen the constraint (*) to the effect that no agent can decide for other agents living in counterfactual worlds. Thus, if $m_1 \in h_a$, $m_2 \in h_d$ and they are contemporaneous ($i_{m_1} = i_{m_2}$) or m_2 belong to a later instant than m_1, then $Choice^{m_1}_{m_2}$ may not be $\{H_{m_2}\}$. In other words, the original person h_a's choice at moment m_2 for the twin h_d at contemporaneous or future moment m_2 can be substantial.

[14] We use singular "they/them" for actual person and "it/it" (singular) or "they/them" (plural) for digital twins and alternative-I's.

Secondly, we require that twins' decisions be more detailed than the original's in the following sense. A substitutional digital twin will not be a totally autonomous or free agent. The original person, as a boss, will have a certain amount of rights to make decisions on their twins' actions, and the twins will be programmed to obey the decisions[15]. On the other hand, there will be some discretion on the part of the twins as to how to implement the decisions made by the original. Imagine the original person ordered a twin to reply an email to refuse the request for peer review of a paper. While the refusal itself is a decision by the original, it is up to the twin to make an excuse for the refusal. Thus, substitutional twins' decisions should be more detailed than the original's. To cash this out, suppose that $m_1 \in h_a$, $m_2 \in h_d$ and $m_2 \leq m_3$. We require that $Choice^{m_2}_{m_3}$ be a *refinement* of $Choice^{m_1}_{m_3}$. That is, every class in $Choice^{m_2}_{m_3}$ should be a subclass of some class in $Choice^{m_1}_{m_3}$. Let us see an example.

In the figure below, we assume that $m_2 = m_3$ for simplicity. So $Choice^{m_2}_{m_2} = \{D_1, D_2, D_3\}$ is a refinement of $Choice^{m_1}_{m_2} = \{C_1, C_2\}$. We have $m \Vdash \Box_{m_1} A$, the fact that A at m is due to h_a's decision at m_1, since A holds at at every moment in the class C_2 to which m belong, and it fails outside of C_2. Then, this entails that $m \Vdash \Box_{m_2} A$, i.e., it is also due to h_d's decision at m_2 since the class D_3 to which m belongs is a subclass of C_2. Thus, a digital twin always obeys its original's decisions[16].



On the other hand, $m \Vdash \Box_{m_2} B$ does not entail $m \Vdash \Box_{m_1} B$ since there is a moment which are choice-equivalent to m but does not force B as shown in the figure.

15 In contrast to this, a dialogical digital twin will behave as a totally autonomous agent, not being subject to control by the original person. Intuitively it should be more profitable to have a dialogue with an autonomous and equal partner than with a subordinate. It is a future task to find constraints on the Alt-I model that represent the dialogical relationship between a person and their digital twin.

16 As mentioned in an earlier footnote, this is the reason we drop the independence constraint for digital twins.

Although the original person decides general policies (e.g., *A*, with a *rough* choice set) as to how a digital twin acts, not all actions made by a digital twin are decided by the original person. The twin tries to put the policies into practice at its own discretion (e.g., *B*, with a *refined* choice set). Now, who will take responsibility if the twins' own decision (*B*) causes a problem? One can see a source of troubles that might be caused when we start using substitutional digital twins.

In sum, here we take the loosening the constraint (*) as characterizing digital twins and counterfactual alternative-I's, and the refinement constraint as describing a power relation between substitutional digital twins and their original person. This is not intended as the completely correct description of how digital twins interact with actual persons. Our aim was to give an application of the notion of alternative-I and the Alt-I model, and a detailed analysis of digital twin is a topic for future research.

4 Conclusion

This paper presented the notion of alternative-I that parametrizes away the paradox of possible-I in the context of decision-making, gave a formal model for the logic of alternative-I based on Belnap's branching-time model, and finally discussed the application of the Alt-I model to digital twin.

Funding This study was partly funded by the academic consulting project, with NTT Service Evolution Laboratories, concerning "the research agendas towards the construction of a new world view harmonizing the developing technology and the human beings," and Interdisciplinary Research Project "From an East Asian view of the self to a society with multi-layered values," Kokoro Research Center, Kyoto University. The third author is partially supported by KAKENHI 19K00022.

Declarations

Competing interests The authors declare no competing interests.

References

Bartha, P. (2014). Decisions in branching time. In T. Müller (Ed.), *Nuel Belnap on Indeterminism and Free Action* (pp. 29–56). Springer.

Belnap, N. D., Perloff, M., & Xu, M. (2001). *Facing the Future: Agents and Choices in Our Indeterminist World* (with contributions by P. Bartha, M. Green, and J. Horty). Oxford University Press.

Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press.
Lewin, K. (1922). *Der Begriff der Genese in Physik, Biologie und Entwicklungsgeschichte: Eine Untersuchung zur vergleichenden Wissenschaftslehre*. Springer.
Lewin, K. (1923). Die zeitliche Geneseordnung. *Zeitschrift für Physik, 13*, 62–81. https://doi.org/10.1007/BF01328203
Lewis, D. K. (1968). Counterpart theory and quantified modal logic. *The Journal of Philosophy, 65*(5), 113–126. https://doi.org/10.2307/2024555
Lewis, D. K. (1986). *On the Plurality of Worlds*. Blackwell Publishers.
Padovani, F. (2013). Genidentity and topology of time: Kurt Lewin and Hans Reichenbach. In N. Milkov, & V. Peckhaus (Ed.), *The Berlin Group and The Philosophy of Logical Empiricism* (pp. 97–122). Springer. https://doi.org/10.1007/978-94-007-5485-0_5
Yagisawa, T. (2017). S4 to 5d. *Argumenta, 2*(2), 241–261. https://doi.org/10.23811/44.arg2017.yag