Branching versus Divergent Possible Worlds
Jiří Beňovský*
University of Fribourg,
Switzerland

Abstract
David Lewis’ modal counterpart theory falls prey to the famous Saul Kripke’s objection, and this is mostly due to his ‘static’ ontology (divergence) of possible worlds. This paper examines a genuinely realist but different, branching ontology of possible worlds and a new definition of the counterpart relation, which attempts to provide us with a better account of de re modality, and to meet satisfactorily Kripke’s claim, while being also ontologically more ‘parsimonious’.

§1) David Lewis’ modal counterpart theory is based on the relation of similarity: in short, something is an other-worldly counterpart of me if it is an individual inhabiting a different possible world who resembles me as for the important features of mine and who resembles me more closely than any other object in that world; it is important that those individuals, my counterparts, are not (numerically) identical to me: they all exist in their own worlds and I am the only one inhabiting the actual world. There are neither causal nor spatiotemporal relations between them and me—what there is, is a relation of similarity and that’s how we can say that they are what I would be, would the world be different.

Counterpart theory leads to a famous difficulty raised by Saul Kripke [2, p. 45] precisely because it is a world-bound individuals theory, which permits to individuals to inhabit just one single world and no more. It is this feature of Lewis’ theory that yields objections showing that it misses the target: counterpart theory does not permit to give a proper account of what may happen to us because what are to be our counterparts are folks which are not related to us in any way, except the fact that they resemble us. So what might happen to me is not about me at all, it is a story about someone else, living in some other world which has nor causal neither spatiotemporal relations with mine. But we have a strong commonsensical belief that we have many properties only contingently and that there are many different ways for us to be different. If, for instance, I just avoided a deadly fall while trying to climb on Mont-Blanc, it is then fundamental, to justify my sensation of relief, that I have the belief that it was me who could have fallen and died. But according to counterpart theory, my idea that I could have died is not really about me: if I say this, I’m speaking about someone else—someone who resembles me and who is my counterpart but who is not me in any way. Following the counterpart theory, we get then a concept of de re modality which is quite different from our commonsensical beliefs about what might happen. Alvin Plantinga [5, p. 164-165] makes the same point in an analogous fashion: if I say being able to perform a certain action A (“It is possible for me to do A.”) then, according to counterpart theory, I’m speaking about someone who does A but who is not me—and of what importance and interest could this possibly be to me? How can actions done by someone else in some other world be of any help to me to know what I can do or not? Indeed, it seems that, according to counterpart theory, I (I myself) can not perform A, for when I say “It is possible for me to do A.” I’m not speaking about myself but about some other individual.

§2) It is not the purpose of this paper to examine the reply to this objection given by Lewis himself. Rather, I shall now try to present another way to meet Kripke’s claim—a way that Lewis himself rejects—which merits our attention and which avoids the aforementioned objections; thus, this paper shall be a discussion of an alternative theory to Lewis’ which will provide with a different structure of possible worlds and

*For very helpful comments on earlier versions of this paper, I would like to thank Gianfranco Soldati, Kevin Mulligan, Martine Nida-Rümelin, Fabrice Correia, Philipp Keller, and Davor Bodrozic.
another definition of the counterpart relation.

§3) Lewis’ structure of possible worlds is a static one and I mean by this more or less what Lewis himself means when he speaks about divergence [4, p. 206]. According to his modal realist framework of possible worlds, worlds are causally and spatio-temporally disconnected. They don’t overlap: no part of a world is part of another. There are worlds with duplicate initial segments: they are exact reproductions of each other and stay as simple copies up to a point from which only they will diverge and differ. Up to the day I decided to climb on Mont-Blanc, there were many worlds, infinitely many maybe, existing independently but being exactly alike, and diverging only afterwards (in some of them I fell down the mountain and died, in others I survived).

Sometimes, Lewis was accused of defending a ‘crazy ontology’ theory because it postulates that there is much more than what we usually believe there to be; there are many counterparts of Mont-Blanc instead of the only one we believe in, there are many more climbers, and so on. I do not believe that Lewis’ theory is more or less ‘crazy’ here than many other of the ‘ersatzist’ alternatives: Lewis multiplies entities of one and the same kind beyond what we usually believe there to be, while others multiply kinds of entities there are, since possible worlds are said to be abstract objects like sets of propositions or abstract states of affairs. Choosing between multiplicity of things of the same kind (of a kind we commonsensically already believe in anyway) and multiplicity of ontological categories, I prefer the first. Besides, any ersatzists postulate as many entities that play the role of possible worlds as there are Lewisian worlds, so I really don’t think ersatzism is any less costly. However, Lewis’ theory is, I would say, unparsimonious and unrespectful of Ockham’s Razor in another sense: as mentioned, it postulates that there is a great number (an infinity, probably) of possible worlds which, for a very long time (seconds, hours, days, billions of years or more), do not differ in any way and are simple copies or duplicates—if there is waste in Lewis’ ontology, then I think it’s this.

§4) A dynamic structure of possible worlds I shall now examine will precisely set this point right: possible worlds will be said to exist only when their existence is required to give an account of what is possible; no copies or duplicates will be allowed and entities will not be multiplied beyond necessity—no waste of entities will be permitted here. How can this be done?

The answer is: branching. Let us suppose that the world has a beginning and that all possible worlds share it—they all have the same origin. From that moment, and at every moment of time since then, it was possible for the world to evolve in many different ways: a certain particle could have interacted with others in a different way than it actually did and later on, consequently, the Milky Way could have been completely different, maybe Mont-Blanc could have measured 8850m and Mount Everest 4807m or there could have been no Mont-Blanc at all. What is important here is that at any moment of time of the evolution of the world different possibilities were open ahead of it—and those possibilities, those alternative evolutions of the world, determine the existence of possible worlds, for in a branching structure of possible worlds, at any moment, at any crossroad where alternative ways are to be chosen, the world will be said to undergo a fission and this will generate as many possible worlds as there were possibilities of evolution at the considered moment. Thus, according to this theory, when I nearly slipped and fell down the mountain, the world split into (at least) two: one where I died and one where I survived.

So those are the two needed primitives for a dynamic structure of possible worlds: a common origin for all possible worlds and genuine fissions of them. Of course, those fissions will generate at every moment of time a great number—an infinity presumably—of possible worlds, but all those worlds will be said to come into existence only at a moment when their existence is required and useful to give an account of certain possibilities: there is no need here to postulate the same great number of ‘parallel’ static possible worlds which do not differ (up to a certain moment of time). It is in this sense that the structure of possible worlds I’m proposing here is a dynamic one, for, at any
time, it does not postulate more entities than necessary and is then more respectful of Ockham’s Razor. In a static structure of possible worlds, as the one defended by Lewis, there is always the same number of worlds, while the dynamic structure provides us with a number of worlds which increases over time. Schematically, it is possible to represent this structure as in Figure 1. In this figure, each entire branch of the structure that begins at the origin represents a possible world, each single arrow represents a part of one or more possible worlds (a world-stage)—which is a state of a world at a given moment of time, each node of the tree represents a fission, and the direction of the arrows indicates the flow of time.

A possible world is thus defined as an ordered aggregate of world-stages (temporal parts of worlds) that forms a complete branch and that has the common origin as its starting point. An example of such a world is represented here by \( W_3 \). Labels “\( W_1 \)” and “\( W_2 \)” don’t denote a ‘complete’ world but only a world-stage, a world in a state at a given moment of time, each node of the tree represents a fission, and the direction of the arrows indicates the flow of time.

All possibilities thus realised at any moment when a world splits will be said to have the same ontological status—all have the same reality as we claim for the actual world; they are of the same kind. How are we then to distinguish between the actual world and the others? How does the actual world gain what seems to be an ontological privilege? And how do we know that we inhabit the actual world?

To these questions, we know what Lewis replies: “actual” is an indexical term, like “present”, “here” or “I”, which expresses simply the fact that I, who am writing here at this moment, am speaking about the world I inhabit—likewise, inhabitants of other worlds can use the word “actual” to speak about their own world, the one they inhabit. This conception of actuality is quite compatible with the branching structure of possible worlds: I inhabit one single world and no more (I am part of one of the ‘branches’ of the structure of possible worlds) and, whenever I use the word “actual” it denotes this world, exactly as the word “I” denotes me.

§5) One could be tempted to speak here about one big multiverse instead of speaking about a multiplicity of possible worlds: a multiverse could be seen as a set or aggregate of all ‘branches’, which would then contain all universes, and all possible worlds. In short, the theory could seem to be an actualist one: a single big (actual) world containing all other ‘possible’ worlds. This must be avoided because there would be no more sense then to speak about “other” or “possible” worlds
since there would be nothing but one huge world. But the splits, as I propose to consider them, guarantee a multiplicity of worlds: all possible worlds generated by those splits are, after a split occurred, entirely causally and spatiotemporally disconnected and isolated, and are all equally real. If $W_i$ split and generated $W_j$ and $W_k$ then, from that moment, $W_j$ and $W_k$ are parts of different worlds, totally inaccessible one from each other. What one might call a “multiverse” could thus, at best, be a shortcut to speak about the set of all possible worlds.

§6) Let us now see how this framework can be used to give an account of de re modality. The account provided by Lewis, which also claims that all individuals are world-bound, appeals to his counterpart theory and gives the following paraphrases for de re modal statements.

(1) $(a \text{ is possibly } F) \leftrightarrow (a \text{ is } F \text{ or at least one counterpart of } a \text{ is } F)$

(2) $(a \text{ is necessarily } F) \leftrightarrow (a \text{ is } F \text{ and all of } a's \text{ counterparts are } F)$

The core idea of this view, the existence of individuals inhabiting other possible worlds who are not identical with individuals of the actual world but who are related (who correspond) to them in a certain way, is a good one, I believe. The Kripkean difficulties that Lewis’ counterpart theory yields are rooted in the fact that those counterparts are causally and spatiotemporally totally disconnected from us (this is a direct consequence of his ontology of possible worlds) and, consequently, those individuals have, we could say, nothing to do with us. But the basic idea of the existence of some kind of counterparts can perhaps be adapted to the branching ontology of possible worlds in such a way that problems yielded by Lewis’ theory which bother Kripke and Plantinga so much can be avoided, while preserving its utility in the paraphrases (1) and (2). A new definition of the counterpart relation that one could propose here is the following (from now on, I will use the term “L-counterpart” to speak about Lewisian counterparts and the term “B-counterpart” when I will speak about the counterpart relation in a branching structure of possible worlds):

(B-CT) An individual $I_1$ existing in a world-stage $W_i$ is a B-counterpart of an individual $I_2$ existing in a world-stage $W_k$ ($W_i \neq W_k$) iff $I_1$ and $I_2$ have been a numerically identical individual in a world-stage $W_m$ which directly or indirectly generated the world-stages $W_i$ and $W_k$.

The example in Figure 3 illustrates this definition. A circle represents an individual existing in the world-stage represented by the arrow on which the circle is situated—for instance, the individual $I_1$ inhabits the world-stage $W_1$. Let $I_1$ be me and let $W_1$ be the actual world just before I started the second half of the route to the summit of Mont-Blanc. $W_1$ then split into (at least) two different worlds: $W_2$ where I reached the summit and $W_3$ where I did not. The individual $I_1$ (I myself in $W_1$) also split into (at least) two individuals $I_2$ and $I_3$; $I_3$ is thus a B-counterpart of $I_2$ and vice versa in virtue of B-CT, for even if they are, after the split, two distinct individuals inhabiting different worlds and are from that moment causally and spatiotemporally completely isolated from one each other, they have been numerically identical (they have been the very same individual $I_1$) in a world-stage that (directly) generated the two world-stages they inhabit at present ($W_2$ and
W₁). The same holds for W₄ and W₅ which are generated by a split of W₂, where in W₄ the individual I₄ dies in an unavioded fall and is a counterpart of I₅ in W₅ who survives and safely returns home.

§7) Let us turn now to some questions concerning identity and the B-counterpart relation. On the Figure 4, let I₁ be the individual I was just before the accident (before a split occurred), I₂ the individual I am now and I₃ my unlucky B-counterpart.

One could object here that this conception leads to the assertion that I₁ = I₂, and I₁ = I₃, but I₂ ≠ I₃, and thus implies that identity is not transitive. Of course, this is not so because it is neither the case that I₁ = I₂ nor I₁ = I₃. For I₁, I₂ and I₃ are not individuals, they are not people, tables, bottles of beer or whatever we usually consider— they are temporal parts of people, temporal parts of tables, temporal parts of bottles of beer, and so on, exactly as W₁, W₂ and W₃ are temporal parts of worlds and not worlds. I₁, I₂ and I₃ must thus be regarded as individual-stages, which can be parts of one and the same individual but, being different stages, they are not (numerically) identical. According to this four-dimensionalist account, there is an individual Iₐ (I use the same convention as before by underlining Iₐ and not I₁) which has as parts the individual-stages I₁ and I₂, and an individual Iₙ which has as parts the individual-stages I₁ and I₃; likewise Wₐ is a world which has as parts W₁ and W₂. The individuals Iₐ and Iₙ share one of their parts in common (namely, I₁) but that does not make them identical, and the two worlds Wₐ and Wₙ also share one part in common (namely, W₁) but are not identical either; we’ll say here that they overlap. This is why this view is a “partial trans-world identity” view: some distinct individuals (whole four-dimensional space-time worms) inhabiting different possible worlds share parts—they are ‘partially identical’. It is not that identity (trans-world identity) is replaced by identity-at-a-time, for there is no genuine trans-world identity involved in this view; it is the B-counterpart relation that is grounded in identity-at-a-time, in the idea of shared temporal parts, defended by four-dimensionalism.

It is maybe a disadvantage of this view that it requires four-dimensionalism to be true. It would, one could say, be better neutral on this point. I think that in general this is right—it is always better for any theory to leave other controversial claims open. But in this case it just seems inevitable and indeed justified, for there is a theory about persistence across possible worlds and it is not so surprising that it solves persistence through time and persistence through possible worlds in a connected way, since time plays a decisive role in the coming in existence of possible worlds.

A seeming ambiguity: look at figure 4 again. When, in W₁, I say “I” what is the denotation of this term? Am I speaking about the individual I₁ (= I₁ + I₂ + ...) or the individual I₉ (= I₁ + I₃ + ...)? Or if, in W₁, I use the expression “the actual world”, am I speaking about Wₐ or Wₙ? I already noted that, due to occurrences of splits, a world-stage can be part of more than one world; the world-stage W₁ is part of the world Wₐ as much as it is part of the world Wₙ. Furthermore, I proposed in §4) above that the B-counterpart theorist should endorse Lewis’ indexical analysis of actuality: if, in W₁, I use the expression “the actual world” I am speaking about “this world”— the world I happen to inhabit. But which one of the two worlds Wₐ and Wₙ is “this world”? The important point to remember here is that the two worlds Wₐ and Wₙ were, at the time of utterance of “the actual world” (in W₁), just one single world: when I pronounced those words, I was part of two overlapping worlds at once (but only at the time when they were one) and, consequently, if by those
words, I denote the world I inhabit, I am denoting all of the worlds that contain $W_1$ as a part. But keep this in mind: I am a world-bound individual and I cannot be said to inhabit more than one world—it is $I_1$, which is a temporal part of me (my individual-stage “at the time of utterance”), who pronounced those words and who can be said to be part of more than one world (as far as those worlds are one, before a split occurs) and who is ambiguously speaking about two worlds $W_a$ and $W_b$. One could find genuinely absurd to conceive that by a singular term such as “the actual world” one can refer to more than one thing at once: but the strangeness is only rooted here in the fact that we usually never take into account the existence of genuine splits of worlds.

The same holds in the case where I say “I” or “this bottle of beer on the table” in $W_1$—all those singular terms denote a plurality of things which have a common part exactly as $I_a$ and $I_b$ share $I_1$. Hence, the word “I” uttered in $W_1$ by $I_1$ refers to two (or more) different individuals but only at a time when those individuals are one (before a split occurs).

§8) The partial trans-world identity view seems to have this advantage over L-counterpart theory: it seems that the story of my avoided fall is really about me and not someone else in another and isolated world, and that I have good reasons to worry, or to be relieved. Remember Kripke’s claim: if I say that it is possible for me to reach the summit of Mont-Blanc or to fall and die or whatever, then I say that I myself can die or appreciate the view from the summit, and not some L-counterpart of me who perhaps resembles me but is not me in any way and is a different individual. B-counterpart theory seems to give a better account of Kripke’s claim—for there has been just one person who, for instance, could have had an accident and this person is me; at the critical moment when I nearly fell down, the world split into (at least) two possible worlds—in one of them I survived but not in the other. The relief I’m experiencing is therefore entirely justified because I rightly feel that it was me who could have fallen and who could have been engaged in a different ‘branch’, in a different world where I would die.

Of course my B-counterpart who wasn’t as lucky is, after the accident, a completely different person but he was identical to me before the split occurred. The word “me” when I use it after the avoided accident denotes me (the individual who is writing now) and “me” pronounced by my unlucky B-counterpart just before his death denotes him (and so, denotes another individual) but I am identical to the person I was before the accident, in the sense of trans-time identity within a single world provided by four-dimensionalism, and so is my B-counterpart.

So, there is an advantage of B-counterpart theory over L-counterpart theory. But perhaps not all worries have been dissipated: because, after all, B-counterpart theory, as well as L-counterpart theory, is a theory of world-bound individuals. Surely, B-counterparts share parts, and this is why they have good reasons to care for each other, whilst L-counterparts have no such reason at all. But it is also true that, after the accident, after the fission occurred, I have no reason any more to care for my B-counterpart resulting from this fission, since from now on, he is just someone else, and our common past does not change anything to it. Maybe the B-counterpart theorist could defend himself against this I-do-not-care claim by claiming that it is all right not to care about my B-counterpart after the accident: I enjoy a relief, because I was engaged in a branch where I had no accident, and I don’t have to feel sad because someone else wasn’t as lucky. As for claims like “If I did not go to high-school, I would have been happier”, I think the B-counterpart theorist could, intuitively enough, claim that when I make such a claim I am imagining myself at a certain time before going to high-school, I am considering the possibilities open ahead of me, and I am concerned about them, in accordance with B-counterpart theory. So, even if some worries could perhaps persist, there really seems to be an advantage of B-counterpart theory over L-counterpart theory as far as the problem with caring and relief is concerned.

§9) But there is a problem rooted in the fact that all B-counterparts have the same origin. For wouldn’t it sound sensible to say that it is possi-
ble that I could have been a monk living in the 12th century? Or, simply, that I could have been born two years before I actually was? It seems that there is an objection against B-counterpart theory, for there is no way for an individual of the 12th century to be my B-counterpart according to B-CT: if the monk existed between 1115 and 1174, then I (I myself) can by no means share any individual-stage with him—there is no world-stage in which we can be numerically identical. So, according to B-counterpart theory, it is not possible for me to be a monk who lived before I did. In short, it seems that my origin and the moment of my origin must be necessary to me (the same holds for the origin of all possible worlds).

On one point, I think, B-counterpart theory is right—I believe that it is not possible for me to live in the 12th century. Let us consider the two following sentences:

(1) It is possible for me to climb Mont-Blanc.

(2) It is possible for me to be a monk living in the 12th century.

Those two cases are fundamentally different: (1) speaks about what I, the individual who is now writing, can really do or be whilst (2) does not—it seems quite impossible to me that I could possibly be a 12th-century monk (I disregard here the possibility of being a time-traveller or the possibility of being a reincarnated monk). Of course, it is possible that there was a monk living some 850 years ago who resembled me a lot, but it simply wouldn’t be me.

Is L-counterpart theory in a better position to give an account of the possibility of my being born two years before I actually was or my living in the 12th century? I don’t think so: Lewis’ possible worlds are spatiotemporally disconnected so it seems impossible to make comparisons such as “this event in W1 happened before that event in W2” for there is no reference point. The same problem arises in the case of comparative statements about places: is it possible that I could stand two meters from the place I actually occupy? It seems that such a statement doesn’t make any sense if we accept Lewis’ ontology—his worlds are distinct isolated space-times, and so to say that something in W1 is elsewhere than something in W2 is nonsensical; there is no way to speak about a reciprocal position between two individuals inhabiting two different worlds (this objection appears in [1, p. 142-143]). So, if an account of such comparative statements (both, temporal and spatial) is wanted, Lewis is committed to accept L-counterparts of places and moments; but the L-counterpart relation is a relation of similarity and it seems to me hard to see how a moment can resemble another—how a space-time point can resemble another. (Compare: B-counterparts of moments and places seem unproblematic, for it is easy to see that a certain place is a B-counterpart of another iff the two places have been one before a split occurred.)

Still, even if we agreed that the consequence of B-counterpart theory that it is not possible for me to be a monk living in the 12th century is maybe not a non-grata one, the fact remains that the B-counterpart theorist’s claim of necessity of origin is a very strong one—not only could I not be a monk in the 12th century, but it is also impossible for me to be born just one year, or even one hour before I actually was, and this, I think, is really a counter-intuitive claim. (So maybe the assumption should be dropped. Maybe the branching structure of possible worlds should be regarded as infinitely branching to the ‘left’ (backwards in time) as it is to the ‘right’ (towards the future). But there was a reason why a common origin for all possible worlds has been postulated: to grant that they all are B-counterparts of each other (and so they are alternative ways things might have been). If the assumption of a common origin is dropped, some possible worlds would turn out to be entirely spatio-temporally and causally disconnected from others and that would simply amount to the Lewisian structure of possible worlds, and the definition of B-counterpart relation would have to be dropped, too.)

§10) Lewis himself rejects genuine branching because, according to him, such a view is contrary to the common sense idea that we have one single future [4, p. 206-208]. If I wonder about what tomorrow will bring to me, then, if there is
branching, this wondering seems to be nonsense; for if there are thirty-four alternative futures for me then all of them are equally mine, all of them will happen and all of them will happen to me (in the case of his static structure of possible worlds, this is not so: I have but one single future, the rest of them belong to my L-counterparts only). I think the B-counterpart theorist could reply thus: I am an individual and not only an individual-stage. When I use the word “I” it denotes a complete four-dimensional person—but, in the case of the branching structure of possible worlds, ambiguously (see above). But still, even if there is an individual-stage of me who is involved in the constitution of individuals having different futures, there is, for each individual, no more than one future; I stressed this point before when I insisted on the fact that an individual, as a whole four-dimensional being, cannot exist in more than one world (is a world-bound individual). So, since an individual-stage does not have a future at all (for it is just a stage which does not last for long), and since an individual exists in only one world and no more, there is no place for a plurality of futures for anyone. What there is room for in the theory, and this is, of course, highly wanted, is a plurality of possible futures. I (I myself, an individual) have many possible futures, for I have many B-counterparts which will evolve differently from me and will live a different future from mine; L-counterpart theory with a static structure of possible worlds gives the same account: there are also many possible futures, lived by L-counterparts of actual individuals who, themselves, have one and one only.

§11) I think there is an advantage of the dynamic branching structure of possible worlds with B-counterpart theory over Lewis’ static structure of possible worlds with an ontology of totally disconnected worlds and L-counterpart theory based on relation of similarity. For B-counterpart theory provides us with a better account of de re modality in harmony with the commonsensical belief that if something is possible for me, then I (I myself) am involved and I have good reasons to care about it—it really is to me that things could have happened otherwise. We have seen that this advantage still seems to be subject to some worries and so one could say that it turns out not to be an entirely satisfactory view in the end, but I think it is better than L-counterpart theory. But there are the branching theorist’s premises, and those are heavy to carry. For we have seen that the claim of necessity of a common origin for all possible worlds, and for individuals inhabiting them, is problematic; besides, it is controversial whether possible worlds (and, to start with, our world) must have a beginning instead of existing eternally and forever, in the first place. Also, it is counterintuitive to accept genuine splits of worlds and individuals. Further, as Lewis points out [4, p. 209], the branching ontology of possible worlds seems less palatable than Lewisian modal realism, because it complicates the account of how possible worlds are unified by spatio-temporal interrelation. For Lewisian possible worlds are quite simply defined in terms of spatio-temporal relations of their parts (any \( x \) that is spatio-temporally related to any \( y \) is part of the same world as \( y \)). In the case of the branching structure of possible worlds, the account is more complicated because there are spatio-temporal relations even between parts from different worlds: on figure 4, \( I_2 \) and \( I_3 \) have a complex spatio-temporal relation because both are spatio-temporally related to \( I_1 \). The branching theorist would perhaps disagree, for he claims that his possible worlds are, after a fission occurred, spatio-temporally unrelated, and so, I agree with Lewis that the objection here is not deadly and consists mainly in the claim that Lewisian modal realism provides a simpler criterion for the unification of worlds.

So, while there are some genuine benefits of the partial trans-world identity view, they are perhaps outweighed by the view’s heavy ontological commitments.

References


\(^2\)It would be interesting to examine how/if the many-worlds interpretation of quantum mechanics could corroborate this premise.

