Real and hyperreal possibility: infinitesimal probabilities in branching time structures

EARLY DRAFT

Sylvia Wenmackers

August 10, 2012

Abstract

We investigate how branching time (BT) structures can be combined with probability theory. In particular, we consider assigning infinitesimal probabilities—available in non-Archimedean probability theory—to individual histories. We illustrate the proposal by applying it to an infinite sequence of coin tosses. We also demonstrate how the approach works in light of the problems of future contingents and historical counterfactuals. We introduce the concept of ‘hyperreal possibility’ as a new modal notion between ‘impossibility’ and ‘real possibility’.

1 Introduction

Branching time (BT) structures are tools in the tense-logical analysis of the problem of future contingents and of historical counterfactuals.

1.1 Problem of future contingents and BT

The problem of future contingents deals with the question of how to ascribe truth values to statements about future events, which are neither inevitable nor impossible to occur (see Øhrstrøm and Hasle, 2011).

In classical times, this question came to the fore in Aristotle’s paradox of the sea battle. Later scholars have interpreted Aristotle’s position as denying that future contingents have a truth value before the relevant event has happened. In his Master Argument, Diodorus Cronus assumed the necessity
of past events to infer the necessity of future events; Diodorus thereby denied the existence of future contingency and arrived at a deterministic view.

In medieval times, the problem recurred in a theological setting as an apparent incompatibility of (i) human free will and (ii) divine omniscience, which includes foreknowledge of human actions. William of Ockham and other medieval scholars such as Richard of Lavenham rejected the view that past events are necessary, allowing for future contingency (and thus human free will). In the mean time, they held the view that future contingents do have a truth value before the relevant event has happened (allowing divine foreknowledge).

In recent times, Arthur N. Prior reintroduced the problem of tense and time in the context of modern logic. Prior (1957) presented the first version of his tense logic, which is closely related to modal logic. In reaction to this book, Saul Kripke suggested the idea of branching time (BT) in a letter to Prior (see Ploug and Øhrstrøm, 2011).

Prior preferred his ‘Peircean’ system, which assumes the necessity of the past, but which entails a rejection of the principle of the future excluded middle; in this system, any future contingent is either true or false.

Prior also developed a different semantics, the ‘Ockhamist’ system, which does assume the principle of the future excluded middle, but rejects the necessity of the past; this system allows for truth value gaps (indeterminacy) of future contingents. Nuel Belnap holds a view close to Prior’s ‘Ockhamist’ system. Also Richmond H. Thomason’s supervaluationism and John MacFarlane’s relativism are related positions. Although Jan Łukasiewicz’s three-valued logic also allows for truth gaps, it is unlike the ‘Ockhamist’ system in that it accepts the necessity of the past and rejects the principle of the future excluded middle.

Despite its name, Prior’s ‘Ockhamist’ system lacks a central aspect of William of Ockham’s position: it allows for truth gaps, so it lacks the notion of a ‘true future’—unknown to any human, but knowable by an omniscient being. In the context of BT structures, the true future has been given the name the “thin red line” (TRL) by Belnap and Green (1994), who do not endorse it. The TRL is defended by Øhrstrøm (2009) and Malpass and Wawer (2012).

1.2 Problem of historical counterfactuals and BT

The problem of historical counterfactuals deals with the question of how to ascribe truth values to statements about conditionals in the subjunctive mode with an antecedent that is false, but which was a historical possibility (see Placek and Müller, 2007). A historical possibility is a (real) possibility at an instant prior to the current one, which is incompatible with the current
1.3 Goal: combining BT with probability

Probability seems to presuppose a notion of possibility: it can be considered as a function that gives weight to various possibilities. Hence, one could expect a very natural connection between probability theory and modal logic (including tense logic), semi-formally given by:

<table>
<thead>
<tr>
<th>Probability theory</th>
<th>Modal logic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu(\phi)$ &gt; 0 ⇔ $\Diamond \phi$ is true</td>
<td></td>
</tr>
<tr>
<td>$\mu(\phi)$ = 1 ⇔ $\Box \phi$ is false</td>
<td></td>
</tr>
</tbody>
</table>

where $\mu$ is a standard probability function, $\phi$ is a proposition, $\Diamond$ is the modal possibility operator, and $\Box$ is the modal necessity operator. The above implications indeed hold for finite sample spaces, but the right-to-left implication is violated in the infinite case.

In the current paper, we investigate how BT structures (formally introduced in section 2) can be combined with probability. We will consider an alternative to standard probability theory to avoid violating the natural connection between probability and modal notions in the case of tense logic. We also hope that the new approach helps to shed new light on old issues related to future contingents and historical counterfactuals.

BT-based probabilities have also been investigated by Müller (2011), but the approach is limited to (i) associating probability spaces to ‘real possibilities’ in structures with finite branching and (ii) combining such probability spaces. Here, we start from a more global perspective by associating a probability space to the entire set of possibilities (histories) in a BT structure. Due to well-known problems related to infinite sample spaces, it will turn out that classical probability theory is not suitable for the task at hand. If we use non-Archimedean probability (NAP) theory instead (see section 3), which allows for the assignment of infinitesimal probability values, it turns out that (i) our conclusions are consistent with those of Müller (2011), and that (ii) we can analyze a more general class of problems of interest.

To preserve the natural link between probability theory and modal or tense logic, it is necessary to work in a probability theory that obeys the principle of ‘Regularity’ (see e.g., Wenmackers, 2012b, section 8.1): such a theory will only assign probability zero to logically impossible events (represented by the empty set).

---

1 Real possibilities’ are defined in section 2.6; for now it suffices to know that they are partitions of histories at a particular moment.

2 Observe the equivalence of sentential (or propositional) algebra (used in modal logic) and of ($\sigma$-)algebra’s of sets (used in probability theory).
In section 4, we propose a new definition for a modal operator that expresses ‘real possibility’, which is contrasted with an operator for ‘hyperreal possibility’. The first corresponds to a possibility that has a non-infinitesimal probability, whereas the second corresponds to a possibility that has a non-zero, infinitesimal probability.

We indicate some further developments that are left for future work in section 5.

In the body of this paper, we take BT structures and probability theory for granted: we assume that the former is a good approach to represent future possibilities in an indeterministic system and we assume that the latter is a good approach to assign a measure to these possibilities. These assumptions are critically discussed in section 6.

1.4 Infinite sequence of coin tosses

Throughout the paper, we illustrate various aspects of BT structures and their combination with probability theory with a toy example: an infinite sequence of coin tosses. At each toss, the coin may land either heads (↑) or tails (↓). We assume that the coin is tossed for a countably infinite number of times, such that the instants at which it is tossed may be indexed with the elements of \( N = \{ 1, 2, 3, \ldots \} \). We refer to \( N \cup \{ n = 0 \} \) as \( N_0 \): it is helpful to include the singleton \( \{ n = 0 \} \) to indicate the last instant prior to the first toss.

We focus on \( N \) for clarity of presentation. However, it is important to observe that the approach is fully general and would apply just as well to a sequence of coin tosses that extends infinitely long into the past as well as into the future, such that the tosses may be indexed with the elements of \( Z = \{ -3, -2, -1, 0, 1, 2, 3, \ldots \} \) or even with a dense set, such as \( \mathbb{R} \).

In section 4, we will analyze a future contingent concerning the toy example:

- “The coin may land heads on each toss.”

We will also discuss the corresponding historical counterfactual, where we assume that the first four tosses were heads, heads, tails, and heads (↑↑↓↑):

- “If the third toss had been heads, the coin could have landed heads on each toss.”

2 Branching time (BT) structures

In this section, we introduce some concepts and notation that are common in the literature on BT structures. Readers who are familiar with this literature may wish to skip it. Except where indicated otherwise, in this section we follow the definitions as given in Müller (2011).
2.1 Possible moments and ‘earlier–possibly later’ relation

A BT structure consists of a pair \( M = (M, <) \), with \( M \) a non-empty set and \(<\) a relation on \( M \) that is:

**Transitive** \( \forall m_1, m_2, m_3 \in M \ ((m_1 < m_2 \land m_2 < m_3) \Rightarrow (m_1 < m_3)) \),

**Irreflexive** \( \forall m \in M \ (\neg (m < m)) \),

**Backward linear** \( \forall m, m_1, m_2 \in M \ ((m_1 < m \land m_2 < m) \Rightarrow (m_1 \leq m_2 \lor m_2 \leq m_1)) \),

**Historically connected** \( \forall m_1, m_2 \in M, \exists m \in M (m \leq m_1 \land m \leq m_2) \),

where \( \leq \) is defined by:

\[ \forall m_1, m_2 \in M (m_1 \leq m_2 \Leftrightarrow (m_1 < m_2 \lor m_1 = m_2)) \].

Elements of \( M \) are called ‘possible moments’ or ‘possible states of affairs’ and the relation symbol \(<\) can be read as ‘earlier–possibly later’.

Further assumptions:

- As is usual in the literature, we assume that there is no maximal moment:
  \[ \forall m \in M, \exists m' \in M (m < m'). \]

- It is often assumed that there is no minimal moment either:
  \[ \forall m \in M, \exists m' \in M (m' < m). \]

However, in the toy example we do assume a minimal moment, \( m_0 \):

\[ \exists! m_0 \in M, \forall m \in M \setminus \{m_0\} (m_0 < m). \]

This simplifying assumption is not essential to the method that we present, which is fully general in this respect.

- For simplicity, we assume finite branching: each moment has a finite number of branches. (We will consider what it takes to relax this assumption in section 5.)

---

3From transitivity and irreflexivity, it follows that the relation \(<\) is also asymmetric:

\[ \forall m_1, m_2 \in M ((m_1 < m_2) \Rightarrow \neg (m_2 < m_1)) \].
2.2 Possible histories and actual history

A history represents a maximal possible course of events—a possible way the world depicted by \( \langle M, < \rangle \) could develop. \( h \) is a ‘history’ if and only if all of the following hold:

- \( h \in \mathcal{P}(M) \),
- \( h \) is linearly ordered:
  \[ \forall m, m' \in h (m < m' \lor m = m' \lor m > m') \],
- \( h \) is maximal with respect to the above linear order.

Hence, in a history, any two distinct elements are comparable via \(<\).

We refer to the set of all histories as \( \mathcal{H}_{ist} \). Consider a particular moment, \( m \). Consider the set of possible histories which contain this moment, \( H_m \).

Observe that the set of all possible histories in a BT structure with minimal element \( m_0 \) can always be written as \( H_{m_0} \).

2.3 Instants

We may define temporal instants as the equivalence classes of moments, \([m]_\sim \in M\), that are cotemporal and hence incompatible (mutually inconsistent). We introduce two conditions for these equivalence classes (see e.g., Placek, 2011, Appendix):

1. \( \forall m \in M, \forall h \in \mathcal{H}_{ist}, \exists! m' \in M ([m]_\sim \cap h = m') \);
2. \( \forall m_1, m_2 \in M, \forall h_1, h_2 \in \mathcal{H}_{ist} ([m_1]_\sim \cap h_1 = [m_2]_\sim \cap h_2 \iff [m_1]_\sim \cap h_2 = [m_2]_\sim \cap h_1) \).

Call the set of all instants (i.e., the set of all equivalence classes \([m]_\sim\)) \( T \) and the linear ordering on this set \( \prec_T \). In general, \( T \) can be a dense set such as \( \mathbb{R} \). We may now introduce a function, \( t \), that expresses the instant at which a particular moment occurs: \( t(m) \in T \). Following Belnap and Müller (2010), we call the tuple \( \langle M, <, T, \prec_T, t \rangle \) a “branching time with date-times” (BTDT) structure. Observe that

\[ \forall m_1, m_2 \in M (m_1 < m_2 \Rightarrow t(m_1) <_T t(m_2)) , \]

but not vice versa (the moments might be inconsistent and hence incomparable by \(<\).

Since there is no maximal element in \( M \), \( T \) is always an infinite set. In the trivial case, when each moment only has a single branch (i.e., linear time), \( M \) has the same cardinality as \( T \), which is at least countably infinite. In general, however, \( M \) is uncountably infinite. (Observe that even for finite branching and countably infinite \( T \), \( M \) can easily become uncountably infinite.)
2.4 Example: BT structure of an infinite coin toss sequence

In our toy example, the set of instants is discrete and countably infinite; we choose $T = \mathbb{N}_0$ and $<_T = <_{\mathbb{N}_0}$. Each moment has exactly two branches, so the set of moments is uncountably infinite: $M = \{m_i \mid i \in 2^{\mathbb{N}_0}\}$. Each moment, $m_i$, can be characterized as the outcome of all coin tosses up to a particular instant: an initial segment of $\{\uparrow, \downarrow\}^\mathbb{N}$. At an instant $n \in \mathbb{N}_0$, there are $2^n$ moments: $\{m_i \mid i \in \{2^n - 1, \ldots, 2^{n+1} - 2\}\}$.

A BT structure can be represented as a graph. For our toy example, this graph is an infinite complete binary tree: see Fig. 1.

![Graph of a BT structure representing an infinite sequence of coin tosses.](image)

Figure 1: Graph of a BT structure representing an infinite sequence of coin tosses. Assuming that there are countably infinitely many instants (indicated by the horizontal lines), there are uncountably many moments (indicated by the blue dots).

In the toy example, each history can be regarded as an element of $\{\uparrow, \downarrow\}^\mathbb{N}$. In Fig. 2, we indicate two particular histories in this BT structure: $h_1 = \{m_0, m_1, m_3, m_7, \ldots\}$ ($=\uparrow\uparrow\uparrow\uparrow\ldots$) and $h_2 = \{m_0, m_2, m_5, m_{12}, \ldots\}$ ($=\downarrow\downarrow\downarrow\downarrow\ldots$).

The actual history may be thought of as one of all possible histories and can be represented as a thin red line (TRL) (Malpass and Wawer, 2012). See Fig. 3 for a graphical example.

2.5 Prior-Thomason semantics for BT structures

A BT model $\mathcal{M}$ is a BT structure $\langle M, < \rangle$ together with a valuation $V$ assigning extensions to atomic propositions at the point of evaluation, $m/h$ (i.e., a moment $m$ and a history $h$ through it: $m \in h$). For a stand-alone sentence uttered in a given context (formally: at a moment of context $m_C \in M$), the parameter $m$ is initialized as $m = m_C$.

Temporal operators move the moment of evaluation along the current history of evaluation: they change $m$, and keep $h$ fixed.
Figure 2: Example of two particular histories in the BT structure of an infinite sequence of coin tosses.

Figure 3: Representation of the actual history as a thin red line (TRL) in the BT structure of an infinite sequence of coin tosses.

$P$ past tense (“it was the case that”)

$\mathcal{M}, m/h \models P\phi$ iff $\exists m' \in h$ such that $m' < m$ and $\mathcal{M}, m'/h \models \phi$;

$F$ future tense (“it will be the case that”)

$\mathcal{M}, m/h \models P\phi$ iff $\exists m' \in h$ such that $m < m'$ and $\mathcal{M}, m'/h \models \phi$.

Modal operators quantify over the possible futures of the moment of evaluation: they keep $m$ fixed, and change $h$.

$POSS$ future possibility

$\mathcal{M}, m/h \models POSS\phi$ iff $\exists h' \in Hist$ such that $m \in h'$ and $\mathcal{M}, m/h' \models \phi$;

$SETT$ necessity or ‘settledness’
\[ M, m / h \models \text{SETT}\phi \iff \forall h' \in \text{Hist} \text{ such that } m \in h' \text{ and } M, m / h' \models \phi. \]

Observe that: \( \text{POSS} = \neg \text{SETT}\neg. \)

### 2.6 Real possibilities and transitions

Consider a particular moment, \( m. \) Define an equivalence relation on \( H_m \) (recall: this is the set of all possible histories which contain \( m \)), \( \equiv_m \) ("are undivided at \( m \)):

\[ \forall h_1, h_2 \in H_m h_1 \equiv_m h_2 \iff \exists m' \in h_1 \cap h_2 (m < m'). \]

Real possibilities at \( m \) are the members of the partition \( \Pi_m \) of \( H_m \) induced by \( \equiv_m \).

Two histories split at moment \( m, \perp_m \), if \( m \) is their last common moment: \( h_1 \perp_m h_2 \) if and only if \( m \) is maximal in \( h_1 \cap h_2 \), \( m \) is a choice point if and only if \( \Pi_m \) has more than one member (i.e., if and only if there are at least two histories splitting at \( m \)).

A transition is an ordered pair \( (m, H) \), also written as \( tr = m \rightarrow H \), with:

- moment \( m \in M \), an ‘initial’,
- real possibility \( H \in \Pi_m \), an ‘outcome’.

We define \( TR \) as the set of all transitions and \( TR_m \) as the set of all transitions with initial \( m \). A transition is trivial if there is no alternative transition with the same initial. If \( m \) is not a choice point, then \( TR_m = \{ tr \} \), with \( tr = m \rightarrow H_m \) a trivial transition. A set of transitions \( \{ tr_i \mid i \in I \} \), where \( tr_i = m_i \rightarrow H_i \), is consistent if and only if \( \cap_{i \in I} H_i \neq \emptyset \) (i.e., if and only if all outcomes can occur together in the same history).

Both \( \Pi_m \) and \( TR_m \) embody the real possibilities open at \( m \); these possibilities form an exhaustive set of mutually exclusive alternatives.

**Example** In the toy example, each moment has two branches. Hence, each moment is a choice point and the set of real possibility has two members at any moment; one corresponds to possible histories in which the coin lands heads (\( \uparrow \)) at the very next instant, and one corresponds to possible histories in which the coin lands tails (\( \downarrow \)) at the very next instant.

### 3 Introducing probability in BT structures

The question we are concerned with in this paper is how we can combine BT structures with probability theory. We will consider two different forms of probability theory—namely, Kolmogorov’s standard probability theory and non-Archimedean probability (NAP) theory. The axioms of NAP theory are introduced in Wenmackers et al. (2012) and the full theory is given in
Benci et al. (2012). NAP theory uses a field of hyperreal numbers, $\mathbb{R}^*$, rather than the standard real numbers, $\mathbb{R}$, to play the role of probabilities. (See Wenmackers, 2012b, for an introduction to hyperreal numbers.) This allows us to assign infinitesimal probabilities to highly unlikely events. Since Regularity is an axiom of NAP theory, it seems well suited to connect the modal notion of possibility with that of having non-zero probability.

Both types of probability theory require us to fix a ‘sample space’ or ‘universe’, $\Omega$: a non-empty, possibly infinite set of atomic possible outcomes. A probability function, denoted by $\mu$ for the standard function and by $P$ for the NAP function, will assign probability values to members of the ‘event space’: a non-empty collection of subsets of $\Omega$. In the case of NAP, the event space is always equal to the full powerset of $\Omega$, $\mathcal{P}(\Omega)$. In case of Kolmogorov’s theory, the event space is written as $\mathcal{A}$ and may be a $\sigma$-algebra strictly smaller than $\mathcal{P}(\Omega)$ (in the case of infinite $\Omega$).

3.1 Various options to choose $\Omega$

Looking at a BT structure from the perspective of a probabilist, one can make the following initial observation: various instants (recall: these are equivalence classes of mutually incompatible but cotemporal moments, $[m]_\sim$) look similar to different partitions of one and the same sample space. See Fig. 4 for a schematic drawing: each partition $\Omega_n$ corresponds to an instant $t_n$, which is the equivalence class of cotemporal moments $[m_{2^n-1}]_\sim$. This suggests (i) that the probabilities assigned to the moments that belong to the same instant should sum to unity and (2) that we should assign the same probability to a moment as to the set of moments that branch from it at the very next instant. However, each instant contains different elements (moments), so the suggestion that they are ‘partitions of the same set’ cannot be taken literally.

It is fruitful to rephrase the initial observation in terms of partitions of histories that contain the cotemporal moments: in this case, we are indeed dealing with various partitions of the same set, $\mathcal{H}_{ist}$, and it is the case that the probabilities of subsets in each partition (all $H_{m'}$ with $m'$ in $[m]_\sim$ for some $m \in M$) sum to unity.

Let us now approach the problem in a more systematic way to ensure that our solution will cover all aspects of interest. First, we make a list of probabilistic questions concerning BT structures one might be interested in. This will enable us to select the most appropriate sample space and event space.

4Branching diagrams indeed occur in the literature on probability to represent various possibilities even though they do not (necessarily) represent any temporal evolution. For instance, Kelly (1996) introduces Baire space using ‘fans’ (equivalence classes of infinite data streams/possible histories) veering off (branching off) a common handle (observed data/past history).
Q1.a What is the probability of a particular moment (state of affairs)?
\[ P(m) = ? \] with \( m \in M \)

Q1.b What is the probability of an arbitrary combination of moments?
\[ P(X) = ? \] with \( X \in P(M) \)

Q1.c What is the conditional probability of a particular moment, given a prior moment?
\[ P(m \mid m') = ? \] with \( m, m' \in M \) such that \( m' < m \)

Q2.a What is the probability of a particular history?
\[ P(h) = ? \] with \( h \in Hist \)

Q2.b What is the probability of an arbitrary combination of histories?
\[ P(Y) = ? \] with \( Y \in P(Hist) \)

Q3 What is the probability of a real possibility at a moment (i.e., a particular subset of histories)?
\[ P(Z) = ? \] with \( Z \in \Pi_m \) (or equivalently \( Z \in TR_m \))

Q4 Introduce probability in a way that harmonizes with the modal operators, such that:
- \( P(\phi) > 0 \iff \text{POSS} \phi \text{ is true} \)
- \( P(\phi) = 1 \iff \text{SETT} \phi \text{ is true} \)

Recall that these modal operators depend on the context: call this moment \( m_C \).
Q1 suggests using $M$ as the sample space. Since there is no maximal moment, $M$ is always infinite and so will $\Omega$ be; the event space of Kolmogorov’s theory, $\mathcal{A}$, will be smaller than $\mathcal{P}(\Omega)$, thereby not containing all arbitrary combinations of moments. Hence, Q1.b requires the use of NAP theory, where the event space is guaranteed to be equal to $\mathcal{P}(\Omega)$.

Q2 suggests using $Hist$ as the sample space. If there are infinitely many moments in $M$ with at least two branches, $Hist$ is an uncountably infinite set. For similar reasons as before, Q2.b requires the use of NAP theory, thereby guaranteeing the event space to be equal to $\mathcal{P}(\Omega)$.

Q3 is dealt with by Müller (2011), where $\Pi_m$ (or equivalently $TR_m$) is used as the sample space. Observe that, under the assumption of finite branching: $\forall m \in M \Pi_m$ is a finite set.

Q4 is obviously the tense logical version of the general connection sought between modal logic and probability theory (cf. section 1.3). Q4 suggests that we need an event space that is a propositional (or sentential) algebra, ranging over propositions such as $\phi$.

3.2 Uniform choice of sample space: $\Omega = Hist$

It may now appear as though there is no uniform choice of $\Omega$ that will allow us to answer questions Q1–4 simultaneously. However, starting from $\Omega = Hist$ (the obvious choice in light of Q2), it does turn out that we can also deal with Q1, Q3, and Q4.

Regarding Q1 (which relates to the discussion at the beginning of section 3.1), observe that “the probability of a moment, $m$” can be interpreted as “the probability of all histories leading to that moment $m$” (i.e., all histories in $H_m = \{h \in Hist \mid m \in h\}$).

- Q1.a can be rephrased as:
  $P(m) \overset{\text{def}}{=} P(H_m) =? \text{ with } m \in M$;

- Q1.b can be rephrased as:
  $P(X) =? \text{ with } X \in \mathcal{P}(H_m \mid m \in M)$;

- Q1.c can be rephrased as:
  $P(m \mid m') \overset{\text{def}}{=} P(H_m \mid H'_m) =? \text{ with } m, m' \in M \text{ such that } m' < m$.

The conditional probability in Q1.c can be computed from the ratio formula using the absolute probabilities in Q1.a, provided that $P(H'_m) > 0$. Q1.c only deals with the case where $m' < m$. However, observe that:

- $P(m \mid m')$ is 1 if $m \leq m'$ (for then $H_m \supseteq H'_m$);

- $P(m \mid m')$ is 0 if $m$ and $m'$ are incomparable (for then $H_m \cap H'_m = \emptyset$).
Regarding Q3, Müller (2011) investigates the combination of standard probability spaces of the form \( PR_m = (\Omega_m = TR_m, \mathcal{A}_m = \mathcal{P}(\Omega_m), \mu_m) \) for different moments, \( m \). However, it seems like one could avoid this complication, by choosing \( \Omega \) large enough from the start. Indeed, all the \( \Omega_m \)'s are contained in \( \Omega = Hist \) and all the \( \mathcal{A}_m \)'s are contained in \( \mathcal{P}(Hist) \). However, unlike the \( \Omega_m \)'s and \( \mathcal{A}_m \)'s, \( \Omega = Hist \) and \( \mathcal{P}(Hist) \) are infinite sets, which becomes problematic in the context of standard probability theory (see next subsection). Probably, this is the very reason why Müller (2011) focuses on the finite \( TR_m \)'s instead. Moreover, taken in isolation, Q.3 does not require the assignment of probabilities to all of \( \mathcal{P}(Hist) \). In the next section we will see that to answer Q3 for the toy example with an infinite coin toss sequence, it suffices to have probability assignments to \( \mathcal{A} = \mathcal{C}(\Omega) \): the collection of cylindrical events of \( \Omega = Hist \).

Also regarding Q4, \( \Omega = Hist \) can still be used. First, we make explicit the context, by writing: \( P(\phi) \overset{\text{def}}{=} P(\phi \circ m_C) \). Then, we define the latter as follows:

\[
P(\phi \circ m_C) = P \left( \{ h \in H_{m_C} \mid \exists m' \in h \text{ s.t. } m' > m \text{ and } \mathcal{A}, m'/|h \models \phi \} \right).
\]

Alternatively, we may quantify over all real possibilities at moment \( m_C \), instead of all possible histories at moment \( m_C \).

**Example** In the toy example, the uniform choice of \( \Omega = Hist \) amounts to \( \Omega = H_{m_0} = \{\uparrow, \downarrow\}^\mathbb{N} \).

### 3.3 Choice of probability theory

Except for cases in which only a finite number of moments have more than one branch, \( \Omega = Hist \) is an uncountably infinite set. With standard probability theory, the choice of \( \Omega = Hist \) will force us to set \( \forall h \in Hist \mu(\{ h \}) = 0 \), which is problematic for Q1.c and Q4. Moreover, the standard probability measure \( \mu \) cannot be defined on all of \( \mathcal{P}(\Omega) \), which is problematic for Q1.b and Q2.b. Therefore, we suggest to use the framework of NAP theory, rather than standard probability theory.

Observe that Q4 can be regarded as a demand for ‘Regularity’ of the probability function, which is explicitly put into NAP as an axiom. Regularity will also ensure that the ratio formula for conditional probabilities is always defined, except when conditionalizing on the empty set (inconsistency), thereby enabling the handling of Q1.c.

### 3.4 Application to an infinite sequence of coin tosses

We now show how a BT structure of an infinite sequence of coin tosses can be combined with its corresponding NAP function, which is described in Benci et al. (2012, section 5.5).
3.4.1 Standard probability space

Using standard probability theory, we are looking for a probability space \( \langle \Omega, A, \mu \rangle \) to describe an infinite sequence of tosses with a fair coin. A generic infinite sequence of coin tosses is written as \( \omega = (\omega_1, \ldots, \omega_n, \ldots) \) with \( \forall i \in \mathbb{N}, i \in \{\uparrow, \downarrow\} \). Hence, the sample space is \( \Omega = \{\uparrow, \downarrow\}^\mathbb{N} \).

First, we consider a special type of events: events in which exactly \( n \) positions of the infinite sequence are known to be heads (\( \uparrow \)) or tails (\( \downarrow \)). Such events are represented by ‘cylindrical sets’. We define a cylindrical set of co-dimension \( n \) as follows:
\[
C_{(t_1, \ldots, t_n)}^{(i_1, \ldots, i_n)} = \{ \omega \in \Omega \mid i_k = t_k \},
\]
with \( \forall k \in \{1, \ldots, n\} (i_k \in \mathbb{N} \land t_k \in \{\uparrow, \downarrow\}) \). Assuming an equal probability of heads and tails for a single toss (fair coin), the probability of an event is halved for each known position in the sequence. Hence, the probability measure on a generic cylindrical set is:
\[
\mu(C_{(t_1, \ldots, t_n)}^{(i_1, \ldots, i_n)}) = \frac{1}{2^n}.
\]

Using Carathéodory’s theorem, this probability measure \( \mu \) can be extended in a unique way to \( A \), the \( \sigma \)-algebra generated by these cylindrical sets. This completes the description of the three components of a standard probability space \( \langle \Omega, A, \mu \rangle \) for an infinite sequence of tosses with a fair coin.

**Some consequences** Using the notation of BT structures, we have that:

- a generic infinite sequence of coin tosses represents a generic history: \( \omega = h \);
- \( \Omega = \{\uparrow, \downarrow\}^\mathbb{N} = H_{m_0} = \mathcal{H}_{ist} \);
- the \( H_m \)'s correspond to cylindrical events.

This approach has the following immediate consequences:

- each \( H_m \) has a non-zero probability that can be computed using eq. 1;
- each individual history \( \{h\} \) has probability zero;
- likewise, each finite set of histories \( F \) has probability zero;
- hence, the conditional probability \( \mu(\{h\} \mid F) \) is undefined, for any finite set of histories \( F \);
- so, even upon learning that \( h \) is the actual history, one cannot update to \( \mu(\{h\} \mid \{h\}) = 1 \).
• nevertheless, the union of all histories has probability unity;
• also the union of all but a finite number of histories has probability unity;
• moreover, there are combinations of histories with an undefined probability (non-measurable sets).

3.4.2 Non-Archimedean probability (NAP) space

We now look into an alternative description of the probabilities pertaining to an infinite sequence of tosses with a fair coin using NAP theory. This amounts to choosing an appropriate NAP space \( \langle \Omega, w, J \rangle \).

In standard probability theory, the limit operation and the range of the probability function are fixed in advance; they are the standard limit of classical calculus and the unit interval of the standard reals, \([0,1]_\mathbb{R}\), respectively. In a non-Archimedean setting, however, one has to adjust the properties of the non-standard limit operation and the range depending on the details of the application—most notably on the cardinality of the sample space of interest. Because this approach is less known, we explain it in a detailed way.

**General recipe** Along the way towards defining the NAP space \( \langle \Omega, w, J \rangle \), we will have to fix eleven ingredients:

1. **Sample space, \( \Omega \)**: a non-empty set of atomic events;
2. **Weight function, \( w \)**: a strictly positive, real-valued, 1-place function on the elements of \( \Omega \), which determines the relative weight of any two atomic events;
3. **Additive measure, \( m \)**: a positive, real-valued, 1-place function on the finite subsets of \( \Omega \), which determines the relative weight of any two elementary events;
4. **Elementary relative probability, \( p \)**: a positive, real-valued, 2-place function, which determines the probability of any event conditional on any elementary event;
5. **Directed set on \( \Omega \), \( \Lambda \)**: the generic choice is \( \Lambda = \mathcal{P}_{\text{fin}}(\Omega) \setminus \emptyset \), but by choosing a smaller \( \Lambda \), additional properties can be obtained for \( P \); this step is crucial (see below for an example);

---

5Although the details are usually not spelled out this way, also in applications of standard probability theory one has to define (1–4), after which the properties of classical calculus together with Carathéodory’s theorem lead straight to the absolute probability function, \( \mu \), in (11).

6By an ‘elementary event’ we mean any event that is represented by a finite set.
(6) **ideal on the p.o. ring** \( \mathfrak{F}(\Lambda, \mathbb{R}) \), \( \mathcal{I}_0 \): informally, an ideal is a set of elements that are negligible in some sense; this informal characterization is made precise as the set of functions that are eventually zero (where ‘eventually’ has to be interpreted according to the partial order on \( \Lambda \)):

\[
\mathcal{I}_0 = \{ \phi \in \mathfrak{F}(\Lambda, \mathbb{R}) \mid \exists A_0 \in \Lambda, \forall A \supseteq A_0 : \phi(A) = 0 \}
\]

(7) **maximal ideal**, \( \mathcal{I}_{\text{max}} \supset \mathcal{I}_0 \): using Krull’s theorem, which is based on Zorn’s lemma, the ideal \( \mathcal{I}_0 \) is extended to a maximal ideal;

(8) **equivalence relation on the ring** \( \mathfrak{F}(\Lambda, \mathbb{R}) \), \( \sim_{\text{max}} \): the idea is that two functions are defined to be equivalent if they differ by at most a negligible amount (where ‘negligible’ is to be understood as an element of the maximal ideal); Formally:

\[
\forall \phi, \psi \in \mathfrak{F}(\Lambda, \mathbb{R}) : \phi \sim_{\text{max}} \psi \iff \exists \varepsilon \in \mathcal{I}_{\text{max}} : \phi + \varepsilon = \psi;
\]

this equivalence relation leads to the following equivalence classes:

\[
[\phi]_{\text{max}} = \{ \psi \in \mathfrak{F}(\Lambda, \mathbb{R}) \mid \exists \varepsilon \in \mathcal{I}_{\text{max}} : \phi + \varepsilon = \psi \}
\]

(9) **set of all equivalence classes**, \( \mathcal{R}_{\Omega, \text{max}} \): this is the set defined as \( \mathfrak{F}(\Lambda, \mathbb{R}) \) modulo \( \mathcal{I}_{\text{max}} \), which forms an ordered, non-Archimedean field;

(10) **algebra homomorphism** on \( \mathfrak{F}(\Lambda, \mathbb{R}) \), \( J \): this is a type of direct limit— we also call it a non-Archimedean limit—defined as follows:

\[
\forall \phi \in \mathfrak{F}(\Lambda, \mathbb{R}) , J(\phi)[\phi]_{\text{max}} \in \mathcal{R}_\Omega ;
\]

(11) **absolute probability**, \( P \): \( \forall A \in \mathcal{P}(\Omega) , P(A) = J(p(A \mid \cdot)) \).

**Applying the recipe to our example** We now specify the appropriate choices to be made in the general recipe above to apply it to an infinite sequence of coin tosses.

1. As before, a generic infinite sequence of coin tosses is written as \( \omega = (\omega_1, \ldots, \omega_n, \ldots) \) with \( \forall i \in \mathbb{N} \in \{\uparrow, \downarrow\} \). Again, we have the sample space \( \Omega = \{\uparrow, \downarrow\}^\mathbb{N} \).

2–4 The fairness assumption implies that we have to assign the same probability to each individual history (or atomic event). Hence, \( w \equiv 1 \). There is no further freedom in the construction of \( m \) and \( p \).

5. We do have some freedom in the choice of a directed set on \( \Omega \), \( \Lambda \). Because of the fairness assumption, we want to construct the NAP space in a way such that:

\[
\forall F \in \mathcal{P}_{\text{fin}}(\Omega) \setminus \emptyset , \forall A \in \mathcal{P}(\Omega) , P(A \mid F) = \frac{\#(A \cap F)}{\#(F)} .
\]
To achieve this, we will have to choose \( \Lambda \) smaller than \( \mathcal{P}_{\text{fin}}(\Omega) \setminus \emptyset \). In particular, we start by focusing on cylindrical events that specify the initial \( n \) tosses: \( i_1 = 1, \ldots, i_n = n \). Such an event takes the form: \( C_{(\alpha_1, \ldots, \alpha_n)}^{(\Omega)} = \alpha \otimes \beta \), with \( \alpha \in \{\uparrow, \downarrow\}^n \), \( \beta \in \{\uparrow, \downarrow\}^N \), and \( \otimes \) stands for the concatenation operation on sequences. Now, we define families of finite sets of such events, \( \lambda_{n,F} \), which are special subsets of \( \Omega \):

\[
\forall n \in \mathbb{N}, \forall F \in \mathcal{P}_{\text{fin}}(\Omega) \setminus \emptyset, \lambda_{n,F} = \{\alpha \otimes \beta \mid \alpha \in \{\uparrow, \downarrow\}^n \land \beta \in F\}.
\]

We define the collection of all these sets, \( \Lambda_{CT} \):

\[
\Lambda_{CT} \overset{\text{def}}{=} \{\lambda_{n,F} \mid n \in \mathbb{N} \land F \in \mathcal{P}_{\text{fin}}(\Omega) \setminus \emptyset\}.
\]

Observe that:

\[
\forall n_1, n_2 \in \mathbb{N}, \forall F_1, F_2 \in \mathcal{P}_{\text{fin}}(\Omega) \setminus \emptyset, \exists F_3 \in \mathcal{P}_{\text{fin}}(\Omega) \setminus \emptyset
\]

\[
(\lambda_{n_1,F_1} \cup \lambda_{n_2,F_2} \subset \lambda_{\max\{n_1,n_2\},F_3}),
\]

which establishes that \( \Lambda_{CT} \) forms a directed set.

(6–7) Now we define an ideal on \( \Lambda_{CT} \), which we extend to a maximal ideal, \( I_\Lambda \). Since this step relies on Zorn’s lemma, the maximal ideal is not unique.

(8–11) The ordered, non-Archimedean field \( \mathcal{R}_{\Omega, J} \) is now determined and so is the algebra homomorphism on \( \mathcal{F}(\Lambda, \mathbb{R}) \), \( J \). This determines the absolute probability function, \( P \) and completes the construction of a NAP space \( \langle \Omega, w, J \rangle \) for an infinite sequence of fair coin tosses.

**Some consequences** In cases of fair probability distributions, it is helpful to state the results in terms of a ‘numerosity’ function, \( \text{num} \). \( \text{num} \) can be regarded as a way of ‘counting’ infinite sets, which is different from cardinality. It is based on the part-whole principle (a strict subset has strictly smaller numerosity, but not necessarily smaller cardinality, than its superset) rather than one-to-one correspondence (equal numerosity implies the existence of a one-to-one mapping, but the reverse implication does not hold in general as it does for cardinality). Although \( \text{num} \) can be introduced axiomatically (Benci and Di Nasso, 2003), it can also be obtained by the non-Archimedean limit of a finite counting function (Benci et al., 2012b) (see also Wenmackers, 2012b, section 7).

Using the NAP space constructed above and \( \text{num} \), we get:

- \( \forall h \in \Omega, P(\{h\}) = \frac{1}{\text{num}(2^n)}, \) which implies \( \forall m \in M \left( P(H_m) = \mu(H_m) \right) \);
- \( \forall H \in \mathcal{P}(\Omega), P(H) = \frac{\text{num}(H)}{\text{num}(2^n)}; \)
- \( \forall H_1 \in \mathcal{P}(\Omega), \forall H \in \mathcal{P}(\Omega) \setminus \emptyset, P(H_1 \mid H_2) = \frac{\text{num}(H_1 \cap H_2)}{\text{num}(H_2)}. \)
Moreover, the above choice of \( J_\Lambda \) ensures that (for the proof, see Benci et al. (2012, section 5.5)):

- \( P(C^{(t_1,\ldots,t_n)}_{(i_1,\ldots,i_n)}) = \frac{1}{2^n} \);
- as a result, for each \( \mu \)-measureable event \( E \), \( \mu(E) \) and \( P(E) \) differ at most by an infinitesimal.

4 Real versus hyperreal possibility

We now return to the future contingents and historical counterfactuals concerning an infinite sequence of coin tosses as introduced in section 1.4. We evaluate whether the application of NAP theory to the corresponding BT structure can teach us anything about the truth values of these propositions. We use this example to motivate the introduction of a new definition of a modal operator that expresses ‘real possibility’, which is contrasted with an operator for ‘hyperreal possibility’. The first is intended to correspond with a possibility that has a non-infinitesimal probability, whereas the second is intended to correspond with a possibility that has a non-zero, infinitesimal probability. The definitions do indeed establish this natural connection between modality and probability in cases like an infinite sequence of coin tosses. However, there are other cases in which the proposed definitions do not have the intended effect.

4.1 Future contingents concerning an infinite sequence of coin tosses

Let us look at the first future contingent: “The coin may land heads on each toss.” From the analysis with NAP, we have established that \( P(\{h\}) = \frac{1}{\text{num}(2^n)} \) for any possible history, \( h \). So, in particular: \( P(h_{\uparrow\uparrow\uparrow\uparrow}) = \frac{1}{\text{num}(2^n)} \).

The denominator, \( \text{num}(2^n) \), is an infinite hyperreal number, so this probability is an infinitesimal. In other words, we assign a non-zero, infinitesimal probability to the coin landing heads on each toss. To assess which truth value we should assign to the corresponding future contingent, we have to determine the strength of ‘may’. In particular, we can introduce two levels of possibility: ‘real possibility’ (\( \Diamond_r \)) and ‘hyperreal possibility’ (\( \Diamond_h \)), which we want to relate to non-Archimedean probability in the following semi-formal way:

\[
\begin{align*}
\text{POSS} \phi \text{ is true } &\iff P(\phi) > 0, \\
\Diamond_r \phi \text{ is true } &\iff st(P(\phi)) > 0, \\
\Diamond_h \phi \text{ is true } &\iff P(\phi) \text{ is a non-zero infinitesimal},
\end{align*}
\]

where \( st \) is the standard part function, which maps a hyperreal number to the unique closest real number (cf. Wenmackers, 2012b, p. 5). Observe
that $\Diamond \phi$ is true if and only if $(\text{POSS} \phi \land \neg \Diamond \Diamond \phi)$ is true. Any standard probability—where it is defined—is equal to the standard part of the non-Archimedean probability: with standard, real-valued probability, we can define real possibilities (as in Placek and Müller, 2007), but we cannot distinguish some (hyperreal) possibilities from impossibility.

In cases like the coin toss example, we can obtain the above relation between NAP theory and modal logic if we introduce the following semantics:

**POSS future possibility**
\[ \mathcal{M}, m_C/h \models \text{POSS} \phi \iff \exists H \subseteq H_{m_C}, \forall h' \in H, \mathcal{M}, m_C/h' \models \phi; \]

**\Diamond real possibility**
\[ \mathcal{M}, m_C/h \models \Diamond \phi \iff \exists \text{infinite} \ H \subseteq H_{m_C}, \forall h' \in H, \mathcal{M}, m_C/h' \models \phi; \]

**\Diamond hyperreal possibility**
\[ \mathcal{M}, m_C/h \models \Diamond h \phi \iff \text{POSS} \phi \land \neg \Diamond \Diamond \phi \]
\[ \text{iff (} \exists H \subseteq H_{m_C}, \forall h' \in H, \mathcal{M}, m_C/h' \models \phi \text{ and } \neg \exists \text{infinite} \ H \subseteq H_{m_C}, \forall h' \in H, \mathcal{M}, m_C/h' \models \phi). \]

Observe that the current definition of future possibility is equivalent to the earlier definition, but that our definition of real possibility is slightly more inclusive than that of Müller (2011): the previous definition included a condition on all the histories in $H_{m_C}$, whereas $H_{m_C}$ minus some finite subset suffices to satisfy the current definition.

Using the above semantics and provided that there are no individual histories that carry a non-infinitesimal probability, we obtain the following link between probability and modality:

**POSS future possibility**
\[ \mathcal{M}, m_C/h \models \text{POSS} \phi \iff P(\Omega \phi \mid H_{m_C}) > 0; \]

**\Diamond real possibility**
\[ \mathcal{M}, m_C/h \models \Diamond \phi \iff P(\Omega \phi \mid H_{m_C}) > 0; \]

**\Diamond hyperreal possibility**
\[ \mathcal{M}, m_C/h \models \Diamond h \phi \iff P(\Omega \phi \mid H_{m_C}) \text{ is a non-zero infinitesimal}; \]

where we assume that the set $\Omega \phi$ in the algebra of sets $\mathcal{H}ist$ corresponds to the proposition $\phi$ in the relevant propositional algebra.

In our first future contingent, the relevant probability is an infinitesimal, so $P(\phi) > 0$ but $st(P(\phi)) = 0$: $\text{POSS} \phi$ and $\Diamond \Diamond \phi$ come out as true, but $\Diamond \phi$ does not.\(^8\) I do not offer a definite position on whether ‘may’ in natural language corresponds more closely to the operator $\text{POSS}$, $\Diamond \phi$, or $\Diamond h \phi$, but I do think that we do need all three operators to represent distinctions that are

\(^8\)We assume that in the example $m_C = m_0$, so the conditionalizing event is the entire sample space ($H_{m_C} = H_{m_0} = \Omega$) and the conditional probabilities reduce to absolute ones.
made in spontaneous discourse. ◁ stresses the existence of a possibility, no matter how improbable, whereas ◇ preselects possibilities that are infinitely more substantive in probability.\footnote{One might even speculate that the dominance of standard probability theory has made us less sensitive to the notion of possibility in the sense of ◁.}

### 4.1.1 Counterexamples

Although the definitions for POSS, ◇, and ◁ as proposed in the previous subsection do establish the desired connection between (non-Archimedean) probability and modality for the case of an infinite sequence of coin tosses, they may fail to do so in other cases.

The current definition of ‘hyperreal possibility’ is based on the existence of only a finite set of histories that make the relevant proposition come out as true. This suffices to describe an infinite sequence of coin tosses and similar examples. However, distinguishing between finite and infinite sets of possible histories does not suffice when a non-infinitesimal probability is assigned to at least one individual history in the BT structure; this can happen if there are branches that contain only finitely many choice points.

Consider the example of a sequence of coin tosses that is interrupted as soon as the coin lands tails for the first time and that is continued otherwise. The graph of the BT structure is depicted in Fig. 5. In this example, there are only countably many choice points. Only the history that represents the coin landing heads at each instant, $h↑↑↑↑\ldots$, has infinitely many choice points and NAP assigns an infinitesimal probability of $\frac{1}{\text{num}(2\mathbb{N})}$ to this history, as before. The other histories, $h_n$, represent a coin that lands heads on the first instants $1\ldots n−1$ and tails on the $n^{th}$ instant; they are assigned the non-infinitesimal probability $P(h_n)=\frac{1}{2^n}$.

By considering this example, we could revise our proposed definitions of ◇ and ◁ in the following way (keeping the definition for POSS as before):

\begin{align*}
\text{◇'} \text{ real possibility} & \quad \mathcal{M},m_C/h=\text{◇'}\phi \iff (\exists \text{ infinite } H \subseteq H_{m_C}, \forall h' \in H, \mathcal{M},m_C/h'=\phi \text{ or } \exists h' \in H_{m_C},(h' \text{ has only finitely many choice points } \land m_C/h'=\phi)); \\
\text{◇'} \text{ hyperreal possibility} & \quad \mathcal{M},m_C/h=\text{◇'}\phi \iff \mathcal{M},m_C/h=(\text{POSS}\phi \land \neg\text{◇'}\phi).
\end{align*}

However, also this revised definition will not guarantee the desired links (between real possibility and non-infinitesimal probability and between hyperreal possibility and non-zero infinitesimal probability) in the most general case. The reason behind this failure is that there are hyperreal numbers of infinitely many orders of magnitude.\footnote{With NAP, we can determine the probability of the union of all $h_n$ with $n \in \mathbb{N}$: $\sum_{n \in \mathbb{N}} P(h_n) = \sum_{n \in \mathbb{N}} \frac{1}{2^n} = 1 - \frac{1}{\text{num}(2\mathbb{N})}$. Hence, the probability of all possible histories, including $h↑↑↑↑\ldots$, is unity—as it should be.}
Figure 5: Graph of a BT structure representing a sequence of coin tosses that stops as soon as the coin lands heads for the first time. In this BT structure, there are less moments than in the BT structure of an uninterrupted sequence of coin tosses and not all of the remaining moments are choice points.

The first way to satisfy the above condition for a real possibility involves an infinite number of branches. However, the probability assigned to a countably infinite subset in an uncountably infinite sample space is usually an infinitesimal. What is of essence here is not just the distinction between finite and infinite, but rather the more general condition that there exists no ‘cardinality gap’ between $H$ and $H_mC$.

Similar complications will arise if one wants to generalize the current proposal from finite to infinite branching. At this point, we assume that these problems are eliminable and that a general semantics can be given. We leave the details of this proposal unfinished and continue with our discussion of the central example, for which the first definitions of the new operators do have the intended effect.

### 4.2 Historical counterfactuals concerning an infinite sequence of coin tosses

We now assume that the first four tosses were heads, heads, tails, and heads (i.e., $h_{\text{TTRL}} \in H_{11\uparrow\downarrow} = H_{m_{17}}$, cf. Fig. 3) and we consider this historical counterfactual: “If the third toss had been heads, the coin could have landed heads on each toss.”

The moment of utterance is $m_{17}$, but the antecedent of the conditional moves the moment of evaluation to the (counterfactual) moment at a prior instant: $m_7$. The set of possible histories containing this moment is $H_{m_7}$, which will play the role of the conditionalizing event in the probabilistic
analysis.\[^{11}\]

We now evaluate the probability of the coin landing heads on each instant, represented by the set \(\{h_{\uparrow\uparrow\uparrow\uparrow\ldots}\}\), given the conditionalizing event, \(H_{m_7}\), is: 

\[
P(\{h_{\uparrow\uparrow\uparrow\uparrow\ldots}\} \mid H_{m_7}) = \frac{2^3}{\text{num}(2^{m_7})},
\]

a non-zero infinitesimal.

Our conclusion concerning the truth value is analogous to that for the corresponding future contingent “The coin may land heads on each toss”: also this historical counterfactual comes out as true when it is considered as a hyperreal possibility, but it is false when interpreted as a real possibility.

Of course, an infinite sequence of coin tosses is a highly idealized concept—it is after all a toy example. In practice we may encounter very long but finite sequences of contingent events, such as a sequence of 100 000 coin tosses. Although the theory developed so far was intended explicitly to deal with infinite sample spaces, we may expect the same qualitative picture to emerge for large but finite sample spaces: one would still like to distinguish between possible but highly unlikely events and possible events that have a more substantive probability. This is indeed possible and the method will be sketched in section 5.4.

5 Extensions and future work

We have already mentioned that the definitions for the new operators \(\Diamond\) and \(\Box\) do not yet establish the desired link between modality and (non-Archimedean) probability in the most general case. Future work will include a crucial reexamination of the conditions in these definitions. Apart from that, we consider the following extensions and generalizations.

5.1 Extension to include BST structures, infinite branching, and DTR

So far, we worked with a simple BT structure. Certain applications may need a more sophisticated approach. Here, we evaluate whether such extensions are possible in principle; sorting out the details is left for future work.

Branching space-time (BST) structures are an extension of BS structures. It seems that the approach for combining NAP theory to BS structures can also be extended to BST structures in a natural way.

In BST structures it is natural to allow for infinite branching. So far, we have only looked at the case of finite branching. However, NAP can be applied to dense sets such as \(\mathbb{R}\), and by making the necessary Cartesian

\[^{11}\text{We can interpret the antecedent slightly differently, as everything else being equal except for the outcome of the third toss. Under this interpretation, we should take into account the fact that the fourth toss came out heads (which is assumed to be independent from the outcome of the third toss) by considering the conditionalizing event } H_{m_{15}} \cup H_{m_{17}} \text{ rather than } H_{m_7}. \text{ However, this will not make a crucial difference in the subsequent analysis.}
products the intended result can be achieved. So, there seems to be no essential problem to achieve this.

To make sense of counterfactuals, it is helpful to introduce a double time reference (DTR), as in the BST structures with DTR of Belnap (2002) or the “branching times with date-times” (BTDT) structure of Belnap and Müller (2010) (see section 2.3). Whereas the valuation of a simple BT structure uses one moment-history pair, \( m/h \), as the evaluation point, the valuation of a BTDT structure requires an additional moment, \( m_e \), for evaluation. As such, it can be regarded as a formalization of the idea of MacFarlane (2003), who suggested that future contingents should be evaluated not only relative to the context of utterance, but also relative to the context of assessment. It may be illuminating to track how the evaluation points of DTR relate to the two indices of a conditional probability function.

5.2 What about the TRL?

We can use the concept of the TRL in a non-controversial way: to model agents who learn new information by updating, we consider probabilistic conditionalization on an initial segment of \( h_{TRL} \), i.e., some \( H_m \) with \( m \in h_{TRL} \). For this application, we do not need to suppose that the TRL is fixed before the relevant contingent events have happened—it is only used retrospectively. It seems to me that one can make sense of TRL in light of DTR, too: by using two reference points, one ‘clips off’ some finite portion of the TRL, which is known at the moment of assessment. Within this scope, one can make sense of the TRL concept without need of considering divine foreknowledge or a god’s-eye-view on the totality of possible histories.

5.3 Hyperfinite BT structures

A different option to introduce infinitesimal probabilities is to replace the standard infinite set \( M \) by a hyperfinite set and to apply Nelson’s probability theory to it (Nelson, 1987). This would require the development of hyperfinite versions of BT structures, which are currently not available. I leave it to the BT community to decide whether exploring this option would be worthwhile.

5.4 Real possibilities with incomparable probabilities

If we have a BT structure in which \( M \) is finite, or if we are interested only in a finite portion of an infinite set \( M \), a different but related problem may occur. Consider two BT structures \( M_a = \langle M_a, \prec \rangle \) and \( M_b = \langle M_b, \prec \rangle \) with \( M_a \supset M_b \) such that \( M_a \) includes certain highly unlikely events and \( M_b \) does not. As an example of such a highly unlikely event, we may consider the occurrence of an earthquake at the decisive moment of a penalty in a soccer game in a region that is usually seismically inactive.
Adding the dimension of probability can help us to switch from $M_a$ to $M_b$, for instance by cutting away the branches with a conditional probability that is below a certain threshold. To allow more flexibility, one may cut branches with probabilities that are ‘indistinguishable from zero’ in a given context. This can be modeled by relative analysis (Wenmackers, 2012a), or—to facilitate the incorporation in the usual BT theory—one might consider introducing this indistinguishability relation in the modal logic directly.

6 General reflections

6.1 Model-dependence of probability and (in-)determinism

In this paper, the BT structure was taken for granted. It was then shown that there is a general choice for the sample space such that the probabilities of all events of interest can be calculated. However, it should be remarked that in many practical situations it not clear how to assign probabilities to the various possible outcomes and—even worse—not all possible outcomes can be foreseen. This situation of ‘profound uncertainty’ is very common, but it is not always clear that one is in such a situation. Therefore, it is worth stressing that all probabilities, even absolute ones, are model dependent. In particular, if the BT structure is missing branches or is otherwise misinformed, the probabilities associated to it will come out wrong too.

Moreover, it is not crystal clear—at least to this author—what the BT structure represents. It is supposed to track the temporal evolution of a system of interest (or even of the entire world), but it seems more plausible that it represents an aspect of our knowledge structure.\footnote{A similar issue arises when considering conditional probability: in some contexts, it is suggestive to interpret the conditionalizing event as a later event (as compared to some conditionalizing superevent such as the entire sample space). Some authors (e.g., Gwiazda) have tried to construct probabilistic puzzles or paradoxes by combining (special) relativity with conditional probabilities. However, it seems that these puzzles disappear upon a critical examination of the interpretation of conditionalization on nested sets in terms of time.}

Sure enough, many authors claim that BT structures represent possibilities in a truly indeterministic world; they stress that the possibilities involved are not merely epistemological ones (for example Placek and Müller, 2007, p. 75). Nevertheless, the same authors demonstrate their BT models with examples involving coin tosses\footnote{The very first example of a historical counterfactual considered by Placek and Müller (2007) reads: “If this coin had shown heads, I would have won my bet.”}—a canonical example in the context of probability theory too. From the view point of classical physics, however, a coin toss is a simple initial value problem with a unique solution (Diaconis et al., 2007). It is our lack of knowledge of the initial conditions that renders the outcome of a coin toss unknown to us—it is an epistemic matter. Determinism or indeterminism is a property of models of the world and there are many
cases in which deterministic and indeterministic models are empirically indistinguishable (Werndl, 2009) (see also Wenmackers, 2012b, section 9.2). This observation demonstrates that the question of whether the world itself is deterministic or not is a metaphysical issue entirely. In as far as we are after a philosophy that is informed by the empirical sciences, we should not try to find knock-down arguments that demonstrate the existence of true indeterminism in the world.\textsuperscript{14} Rather, it suffices to say that we work within an indeterministic model of the world.

To motivate the use of BT structures and probability theory, it suffices that there are situations in which it is useful to treat coin tosses and the like as indeterministic models. However, because these are models of the world, it is misleading to call them “Our World” (cf. Belnap, 1992).

6.2 Branching structures for the unobserved past?

In the previous subsection it was suggested that branching structures do not track the temporal evolution of a system, but rather our knowledge of this system.\textsuperscript{15} The future is always unobserved by us, but also the majority of past and current events remains unobserved or unremembered by humans. This shows that our lack of knowledge is not limited to the forward direction of time. Since we do have the most information about the recent past, forward branching (with the direction of time) is indeed a plausible representation of our uncertainty concerning future events: we do not know for sure how things will develop, but we foresee a number of possibilities, which in turn allow for various developments. (It is doubtful that we can track this adequately for the far future; hence, the practical application of BT structures seems limited to a rather modest time span—that which we call the foreseeable future. Beyond that we are faced with profound uncertainty.) However, we also lack detailed knowledge of the (far) past. (Otherwise we would not need any research on the subject of history.) Now we may ask ourselves whether we should map our uncertainty concerning the past as forward or as backward branching. If we want to represent the direction of (probabilistic) causation, forward branching seems the best option. However, since we try to “connect the dots” from a limited number of clues about the past (e.g., some fossils, archeological artifacts, and \textsuperscript{14}C dating) to the current state of affairs, backward branching may be a more natural choice to represent our reasoning process.

\textsuperscript{14}It does seem that researchers who work on branching structures are motivated to keep their models in touch with contemporary physical theories. Ever since the second letter of Kripke to Prior (Ploug and Øhrstrøm, 2011) an effort is made to develop branching structures that are compatible with relativity theory (e.g., Belnap, 1992). And often, ‘true possibilities’ are taken to be physical possibilities from quantum theory (e.g., in the context of the Everettian many-worlds picture Belnap and Müller, 2010).

\textsuperscript{15}Observe that also the ‘fanning diagrams’ of Kelly (1996) track an evolution of observed data/increased knowledge, rather than time itself.
6.3 A counterargument against infinitesimal probabilities?

Williamson (2007) has offered an argument against infinitesimal probabilities assigned to an infinite sequence of coin tosses. We can now reconsider this argument in the context of BT structures. In the example, we have presented an infinite sequence in which the individual tosses are indexed by $T = \mathbb{N}_0$. This fixes the present instant (or at least the beginning of observations) at index $t = 0$. If we conditionalize on the initial segment of a history, say up to an instant $t > 0$, the conditional probability of any history that contains this initial segment is a factor $2^{(t + 1)}$ larger than the corresponding absolute probability. Of course, the choice of 0 as the present moment is completely arbitrary. We are free to choose a later index $t > 0$ as the present moment; doing so will yield a different (larger) infinitesimal probability to any history starting from the present moment. In any case, this shows the model-dependence of absolute probabilities. It is a counterargument against infinitesimal probabilities only if one presupposes that such model-dependence does not exist. Indeed, classical probability theory does not violate translation symmetry in this case, but it does assign probability 0 to individual histories, which has its own drawbacks. (For instance, it makes the theory unsuitable to express any learning from conditionalizing on a finite, initial segments.)

Recall that NAP also allows us to model an infinite sequence in which the individual tosses are indexed by $T = \mathbb{Z}$. In this case it may still be suggestive to consider the moment indexed by $t = 0$ as the present.\footnote{Actually, now there are infinitely many, mutually inconsistent moments in the instant $t = 0$!} If the outcome of all past tosses are known, one may conditionalize on this information, after which the probabilities come out exactly the same as in the model where time starts at $t = 0$.

References


