Towards a New Theory of Historical Counterfactuals

JACEK WAWER AND LESZEK WROŃSKI

Abstract: We investigate the semantics of historical counterfactuals in indeterministic contexts. We claim that "plain" and "necessitated" counterfactuals differ in meaning. To substantiate this claim, we propose a new semantic treatment of historical counterfactuals in the *Branching Time* framework. We supplement our semantics with supervaluationist postsemantics, thanks to which we can explain away the intuitions which seem to talk in favor of the identification of "would" with "would necessarily."

Keywords: Conditionals, counterfactuals, modality, branching time, David Lewis, Robert Stalnaker

1 Introduction

There is a recurring idea in the semantics of counterfactuals summarized well by the following statement: "all counterfactual conditionals express necessitation" (Pollock, 1976, p. 34). It was shared by Lewis (1986) for whom, as Stalnaker (1981, p. 93) neatly puts it, "the antecedents of conditionals act like necessity operators on their consequents." Recently, a similar claim was made by Leitgeb (2012, p. 36) who takes

(A) If the match were struck, it would light.

and

(B) If the match were struck, it would necessarily light.

to be synonyms.

We do not find this semantic identification convincing. Not only we do not think that "necessitated" reading of "would" is the default one, but we are even uncertain if there is any non-artificial reading of "would" that has the meaning of "necessarily would."

Indeterministic contexts support our intuition. Consider first a fair coin that I am about to toss and a pair of sentences:

(C) The coin will land heads.

(D) The coin will necessarily land heads.

While we are strongly inclined to reject (D) (on the grounds that the coin is fair), our reaction towards (C) is much more hesitant; we can even say "Who knows, maybe." In terms of degrees of belief, our degree of belief in (C) is much higher (about 0.5) than our degree of belief in (D) (about 0), which itself should suggest that we do not naturally take them to be synonymous. Now, suppose that I could have but did not toss the coin and consider a pair of conditionals:

(E) If I had flipped the coin, it would have landed heads.

(F) If I had flipped the coin, it would have necessarily landed heads.

We claim that our intuitive reactions towards (E) and (F) resemble those towards (C) and (D) respectively. Similarly, our degree of belief in (E) (about 0.5) is much higher than our degree of belief in (F) (about 0). We think that just as (C) and (D) clearly differ in meaning, so do (E) and (F). Observations of this sort support the view that "would" should *not* be be identified with "necessarily would."

To validate our intuition we develop, against the background of the theory of branching time (section 2), a novel semantic treatment of counterfactuals (sections 3 and 4). We combine our semantics with supervaluational postsemantics (section 5) and use this rich conceptual apparatus to explain the subtle affinity of plain and necessitated counterfactuals (section 6).

We choose the branching-time setting since it suits the discussion of indeterminism especially well. This setting has already been used for analysis of counterfactual future contingents (Placek & Müller, 2007; Thomason & Gupta, 1980). However, our semantic approach is, as far as we know, a novel one. We will minimize the discussion of the formal part of the theory to the necessary minimum, leaving the full description of the formalities involved for a different paper.

2 Branching time

Branching time is a natural way of representing how possibilities evolve in time. The framework aims at explicating the idea that some eventualities which are possible at some earlier time cease to be possible at later times. It is also meant to capture the intuition that the future, contrary to the past, is open to multiple possible realizations. Additionally, it incorporates the observation that what is possible and what is settled is very much circumstance-dependent: what is possible at a given moment depends on what the world is like at this very moment (such situation-dependent necessity was sometimes called "accidental necessity" in the scholastic terminology).

All these intuitions are meant to be captured by a pictorial representation of temporal possibilities in the shape of a tree which branches upwards, but never downwards. Each point at a tree represents a possible state of the world – these possible states are called "(possible) moments". At a given moment, each "branch" growing out of a given moment represents its possible future continuation while the "trunk" of the moment represents its unique possible past. Each maximal line throughout the tree represents a full possible course of events – a "(possible) history". ("Moment" and "history" will be used as technical terms from now on.)

Here is an example of a simple branching-time structure used as a depiction of the coin-tossing story we introduced above:

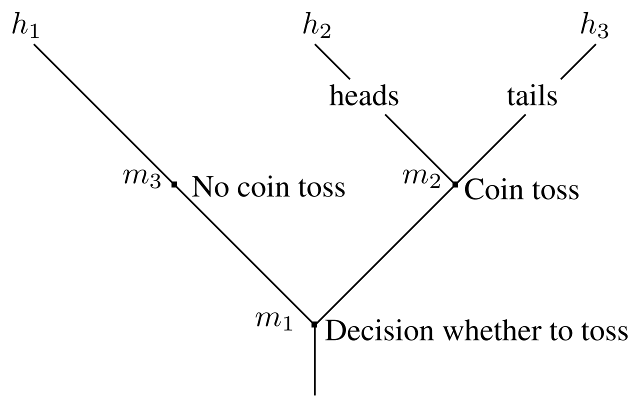

Figure 1: Coin toss

Formally put, a branching time (BT) structure is a partially ordered set (i.e. the ordering relation is reflexive, weakly antisymmetric, and transitive) satisfying the extra condition of no backward branching (or simply backwards-linearity):

- $\forall m, m_1, m_2\big[\big((m_1 < m) \wedge (m_2 < m)\big) \rightarrow \big((m_1 < m_2) \vee (m_2 < m_1) \vee (m_1 = m_2)\big)\big]$.

Elements of the set represent possible moments (and will be called "moments"), and the ordering represents the relation of *earlier-possibly later*. A "history" is a maximal linearly ordered subset of the set.

To give a formal account of our idea, we need an extra notion of "copresence" relation ($\sim$) defined on possible moments. Two moments stand in this relation if and only if they happen at the same time, possibly in different histories. Formally, ($\sim$) is an equivalence relation on possible moments, the standard definition of which can found in (Belnap, Perloff, & Xu, 2001, pp. 194–195).

Branching time structures are meant to interpret the language in which we talk about the future, past, and temporal possibility; it is often used as a model for the propositional language with temporal operators "It will be the case that" (F) and "It has been the case" (P) and a modal operator of historical possibility ($\Diamond$). In the standard Ockhamist semantics, the truth value of a sentence depends on both a moment and a history at which it is evaluated. (Sentences are evaluated at moment-history pairs m/h such that $m \in h$.) For example, the sentence $\mathrm{F}\phi$ is true at a moment-history pair m/h if and only if there is a later moment m' in history h at which ϕ is true. A modal sentence $\Diamond\phi$ is true at moment-history pair m/h if and only if there is a history h' "passing through" m (i.e. min h) such that ϕ is true at m/h'.

3 Semantics of historical counterfactuals

Since we want to use BT structures to model the behavior of counterfactuals, we need to introduce an additional two-argument operator $>$. The basic idea behind the semantic interpretation is rather standard for the possible world setting: the truth value of a counterfactual at a moment-history pair, depends on whether the consequent is true at the "closest"[1] moment-history pair at which the antecedent is true. The BT structure provides a very natural account of "closeness" of possibilities. From the perspective of a moment-history pair m/h, the closest alternative m'/h' is the one that (a) makes the antecedent true; (b) the history h' "branched off" the history h as recently as possible to make the antecedent true; and (c) the moment m' is co-present with the moment m. Thus, to evaluate the counterfactual, we are looking for the history which shares as much past as possible with the actual history; in other words, the history which deviates from the actual history as recently as possible to make the antecedent true now.

[1] Yes, of course there might be ties. Hold on, we will deal with this.

To give a formal account of this idea we follow the lead of Stalnaker (1968) and add a selection function parameter of truth. A selection function s takes as its argument a pair consisting of sentence ϕ and a moment-history pair m/h, while delivering as its value the "closest" moment-history pair m'/h' such that:

1. ϕ is true at m'/h' ($m'/h' \vDash \phi$),
2. $m \sim m'$, i.e. m is co-present with m',
3. $\forall m'' \forall h''((m'' \sim m \wedge (h \cap h' \subset h \cap h'')) \rightarrow m''/h'' \nvDash \phi)$.

For the sake of simplicity, we ignore the possibility of ever-closer ϕ-histories and counterfactuals with historically impossible antecedents (if you require a stance on the latter issue, we suggest you follow the lead of Thomason and Gupta (1980, p. 74) and declare all such counterfactuals as true).

We also ignore, due to the limitation of space, the so-called "late departure" problem and counterfactual future contingents like "If I had bet tails, I would have won" which are notoriously difficult to tackle in the simple BT setting (examples of this sort are discussed in Placek & Müller, 2007; Thomason & Gupta, 1980).

It is important to note that conditions 1–3 above under-determine the choice of a selection function. This is clearly visible in indeterministic contexts. Take the coin toss model above (Figure 1 two pages earlier): the heads-history h_2 branches off the history h_1 at the same moment as the tails-history h_3. Therefore, both these histories are equally viable candidates for values of a selection function for the argument $\langle \textit{coin toss}, m_3, h_1 \rangle$. It is a very important feature of our semantics which we are happy to embrace.

With the selection function at our disposal, we can define the truth conditions of a counterfactual connective $>$:

Definition 1 (Counterfactual) $\quad m/h, s \vDash \phi > \psi$ *iff* $s(\phi, m/h), s \vDash \psi$.

So, a counterfactual is true at a triple *moment-history-selection function* $m/h, s$ iff the selection function s when given the antecedent of the counterfactual and the m/h pair as its argument gives back a *moment-history* pair at which the consequent is true. Observe that the truth of a sentence is relative to a selection function while the definition of the selection function appeals to the notion of truth. To avoid circularity, we define $\vDash$ and the set of all selection functions bottom-up by use of double induction. The curious reader can consult the rigorous definition stated in Appendix B.

Already at this point we have achieved some interesting goals. Our semantics generates some validities we desire to preserve. For example, the counterfactual law of excluded middle $(A > B) \vee (A > \neg B)$ is valid and what we believe to be the natural interaction of counterfactual with negation is preserved, i.e. $\neg(A > B) \leftrightarrow (A > \neg B)$ is true (unless A is a historical impossibility).

4 Counterfactuals and historical modality

The notion of historical necessity is inherently local. However, introduction of a counterfactual connective indicates "the transition to a tense logic in which what is true at moments co-present with i can be relevant to what is true at i" (Thomason & Gupta, 1980, p. 78). This significantly influences the behavior of historical possibility, especially when the whole counterfactual is inside the scope of the possibility operator. In such cases, to determine what is necessary and what is possible at a given moment, we need to take into account what must and what could have happened at different moments, situated in alternative scenarios. For example, to determine whether "It is possible that if yesterday I had met a person in need, I would have helped" is true, I need to know the possible outcomes of the encounter I could have had the day before.

To take these alternative scenarios into account, we need to modify the behavior of the possibility operator, extending the domain of quantification in its semantics. So far, to determine what is possible, we had to take into account the histories passing through the moment of evaluation only, but as soon as counterfactuals are introduced, we need to somehow consider the counterfactual possibilities. We propose to do it in the following manner:

Definition 2 (Possibility) $\quad m/h, s \vDash \Diamond\phi$ *iff* $\exists_{s'}\exists_{h'} m/h', s' \vDash \phi$.

Thus, $\Diamond\phi$ is true at a moment m, history h passing through that moment and according to the selection function s iff there is a history passing through that moment which makes ϕ true according to some selection selection function s'. (We preserve the natural duality of modalities: $\Box = \neg\Diamond\neg$.)

Thanks to this definition we can semantically distinguish "plain would" counterfactuals from "necessitated would" counterfactuals. The coin toss model from Figure 1 can serve as a paradigm example: on the one hand, at

m_3/h_1 the sentence "If I had tossed the coin it would have landed heads" (coin toss $>$ F heads) is either true or false *depending on the choice* of the selection function.[2] On the other hand, the sentences "If I had tossed the coin it would have been necessary that it would land heads" (coin toss $>$ $\Box$ F heads) and "Necessarily, if I had tossed the coin it would have landed heads" ($\Box$(coin toss $>$ F heads)) are both false *independently* of the choice of the selection function. We will give a more thorough interpretation of this phenomenon in the next section.

Simultaneous introduction of possibility and counterfactual operators to the language opens a research perspective regarding their interactions. The pioneering research in this domain has been conducted by Thomason and Gupta (1980). Our definitions significantly differ and simplify those of our predecessors, nonetheless, we were able to restate some of their results. The conclusions of our preliminary research can be found in appendix C.

5 Semantics and postsemantics

To further explore the behavior of plain and modal counterfactuals we use the conceptual apparatus of John MacFarlane (2003, 2014), who distinguishes truth-at-an-index ($\vDash$) from truth-at-a-context ($\Vdash$). The first notion of truth is largely a technical tool of a given semantic theory. It is used to guarantee that the semantic theory validates intuitive tautologies and rules of inference. It is also meant to guarantee the natural interaction between different logical connectives. Moreover, it is used to explicate the notion of truth-at-a-context. This last notion is more closely related to the ordinary practice of speech. At some points (e.g. 2014, p. 208) MacFarlane even suggests that a sentence's truth-at-a-context is sufficient and necessary for accuracy of an assertion of that sentence in that context. MacFarlane calls the theory of truth-at-an-index "semantics" and the theory of truth-at-a-context "postsemantics". For our simple application, the context might be reduced to the moment on the tree where the sentence is used; in turn, the index reduces to the *moment-history-selection function* triple.

Following Kaplan (1989) and Belnap et al. (2001) we understand the notion of the context as metaphysically loaded (it represents some aspects of the world at the moment of utterance), therefore taking the notion of truth-at-a-context to be metaphysically relevant as well. Specifically, we assume

[2] Perhaps the antecedent in the formal representation of the sentence should include the P operator; we ignore this issue as it is not important for the current task.

that the sentence is true-at-a-context if and only if the context is sufficient to ground the truth of that sentence at that context. We do not believe there is anything in the context of use of the counterfactual in virtue of which the counterfactual is true; at the same token we think that there is nothing in virtue of which the negation of the counterfactual is true. Consequently, we propose a postsemantic theory according to which neither a counterfactual future contingent nor its negation is true at a context at which it is used. It turns out that given the semantic definitions we proposed, it is sufficient to slightly generalize supervaluationism of Thomason (1970) to get the appropriate result.

Definition 3 (Truth-at-a-context)

$$m \Vdash \phi \textit{ iff } \forall_s \forall_h \; m/h, s \vDash \phi.$$

That is, a counterfactual is true-at-a-context only if the choice of a "maximally close" counterfactual scenario (i.e. of a selection function) makes no difference since all the maximally close alternatives make the consequent true. The counterfactual is false-at-a-context if and only if its negation is true-at-a-context. This definition of postsemantics is very close in spirit to the theory of counterfactuals advocated in (Stalnaker, 1981).

Such a postsemantics generates the desired result: Every counterfactual future contingent is neither true nor false at the context of its use. Take again the coin toss model from Figure 1, p. 3 as an example. At the context m_3 neither "If I had tossed the coin, it would have landed heads" nor "If I had tossed the coin, it would not have landed heads"[3] is true:

$$m_3 \nVdash \text{coin-toss} > \text{F heads} \;\;\&\;\; m_3 \nVdash \text{coin-toss} > \neg\text{F heads}.$$

It is so because the truth-at-an-index of each of these counterfactuals crucially depends on the choice of selection function. At the same time, the postsemantics guarantees that the disjunction of the two is true at context m_3:

$$m_3 \Vdash (\text{coin-toss} > \text{F heads}) \vee (\text{coin-toss} > \neg\text{F heads}).$$

Interestingly, $A > B$ and $\Box(A > B)$ are true at the very same contexts, but they might be false at different contexts! This observation brings us to the last section of the paper.

[3] Remember that in our semantics "If I had tossed the coin it would not have landed heads" is equivalent to "It is not the case that if I had tossed the coin it would have landed heads."

6 Giving necessitarians their due

In this part, we try to explain away the lure of the idea that all counterfactuals should be read in the strong, necessitated sense. One of the most distinguished proponents of this idea is David Lewis (1986) who motivates this approach by the observation that to reject a counterfactual like "Had I tossed the coin, it would have landed heads," it is often enough to say "No! If you had tossed the coin, it might have landed tails." (p. 8). Since it is sufficient to use a might-not-counterfactual to deny a would-counterfactual, it is tempting to hold that a "would" has something of a "must" in it. Since we try to resist this temptation, we need to explain away Lewis' observation. We believe (in line with the views of Stalnaker, 1981) that the phenomenon is postsemantic rather than semantic in nature and that it should be explained by the criteria of correctness of assertions rather than by the semantics of counterfactuals. We can utilize the notion of truth-at-a-context to this purpose if we accept the following norm:

Definition 4 (Truth norm of assertion) *An act of assertion is correct in a context only if the sentence asserted is true at that context.*

This norm of assertion together with supervaluational postsemantics naturally generates correctness gaps. Since counterfactual future contingents are true at no contexts, they can never be correctly asserted. For example, neither "If I had tossed the coin, it would have landed heads," nor "If I had tossed the coin, it would not have landed heads" can be correctly asserted. Consequently, one can correctly assert a counterfactual only if it is settled. To show that a counterfactual is not true at a given context it suffices to establish that it might false. Together with the truth norm of assertion, it follows that such a counterfactual cannot be correctly asserted. This fact can be summarized by the following postsemantic theorem:

$$m \not\Vdash \psi > \phi \ \text{ iff } \ m \Vdash \Diamond(\psi > \neg\phi).$$

Therefore, it is enough to establish possibility of falsity of a counterfactual to undermine the right to assert it. It is the reason why we can deny "If I had tossed the coin, it would have landed heads" with "If I had tossed the coin, it might not have landed heads."

This postsemantic fact also explains why it is never correct to assert "If I had tossed the coin it would have landed heads, but if I had tossed it, it might not have landed heads." The reason is not that the sentence is self-contradictory, as Lewis would have us believe, but because the sentence is true in no context and as such can never be correctly asserted; that is, for any m in any model:

$$m \nVdash (\phi > \psi) \wedge \Diamond(\phi > \neg\psi).$$

Therefore, the intuition that "would" has a necessitating force is explained by a general linguistic mechanism (truth norm of assertion) rather than by the specific modal strengthening of the meaning of "would". Observe that the very same mechanism is at play in different kind of contexts. For example: "I were at the party, but possibly I weren't" sounds just as strange, but we do not conclude that the sentence is self-contradictory.

We noted in Section 4 that plain and necessitated counterfactuals can be semantically distinguished. This fact has a postsemantic consequence: even though would- and would-necessarily-counterfactuals are true at the very same contexts, they are false at different contexts (i.e. their negations are true at different contexts); that is, there is a model with a moment m such that:

$$m \Vdash \neg\Box(\phi > \psi) \text{ and } m \nVdash \neg(\phi > \psi).$$

Any indeterministic example can attest to that. Let us get back to the coin tossing example. The sentence, "If I had tossed the coin, it would necessarily have landed heads," is false, while the sentence, "If I had tossed the coin, it would have landed heads," is neither true nor false. We think that it follows the linguistic intuition rather well. It also suggests that the semantic identification of "would" and "necessarily-would" is not justified.

Lastly, let us come back to the issue of the relationship between counterfactuals like (A) and (B) (p. 1). Again, we see a postsemantic rather than semantic connection. There are two ways of describing our point.

First, it is sometimes maintained (see e.g. Stalnaker, 1981) that counterfactuals whose natural language formulation includes a word like "necessarily" in the consequent have the logical form in which the corresponding modal operator stands in front of the whole formula, so e.g. "If the match were struck, it would necessarily light" has the logical form of $\Box(A > B)$. If we take this route, notice that in our framework the following rules hold:

$$\frac{m \Vdash (\phi > \psi)}{m \Vdash \Box(\phi > \psi)} \qquad \frac{m \Vdash \Box(\phi > \psi)}{m \Vdash \phi > \psi}$$

Second, if we insist that the aforementioned sentence has the logical form of $A > \Box B$, then the issue is more subtle, but the connection of the two is still close enough. The sentence $\phi > \psi$ does follow from $\phi > \Box\psi$ and the converse entailment holds under an assumption (satisfied by most everyday counterfactuals, including those used in Leitgeb's example) that the antecedent of the counterfactual is not a future contingent. We can express these two observations in the form of the following rules:

$$\frac{m \Vdash (\phi > \psi) \wedge (\phi > \Box\phi)}{m \Vdash \phi > \Box\psi} \qquad \frac{m \Vdash \phi > \Box\psi}{m \Vdash \phi > \psi}$$

The bottom line is that sentences used in Leitgeb's examples – counterfactuals and necessitated counterfactuals – are not synonymous, but their affirmations are assertible in exactly the same contexts, which explains why they look like synonyms.

A Acknowledgments

We would like to thank Tomasz Placek, Thomas Müller, Michael De, Antje Rumberg, the audiences of the ∃ntia et Nomin∀ and Logica 2014 conferences, as well as members of the Izydora Dąmbska Research Group for feedback.

The part of the project realized by Jacek Wawer has been financed by means of the Polish National Science Center granted by the decision number DEC-2013/11/N/HS1/04805. Jacek Wawer also gratefully acknowledges the financial support of the Foundation for Polish Science received in years 2014–2015 in form of the scholarship START no. 122.2014.

B The details of the truth definition

Introductory comments:

- Assume M/H is the set of all moment-history pairs of a given *BT* model such that $m \in h$; when we write $m/h \sim m'/h'$ we mean "m is co-present with m'".
- Assume the issue of everywhere false antecedents is dealt with in some suitable way; e.g. declare that all such counterfactuals are true.
- Assume the standard definition of Ockhamist truth $\vDash_{BT}$.

Definition 5 (Alphabet) *The alphabet of our language consists of:*

- *countably many propositional variables;*
- *parentheses;*
- *connectives:* $\neg$, $\wedge$, $\Diamond$, F, P, $>$.

Definition 6 (Formula)

1. *Every propositional variable is a formula;*
2. *If* ϕ, ψ *are formulas, then* $\neg\phi$, $\phi \wedge \psi$, $\mathrm{F}\phi$, $\mathrm{P}\phi$, $\Diamond\phi$, $\phi > \psi$ *are;*
3. *Nothing else is a formula.*

The class off all formulas is denoted by $\mathcal{F}$*; the class of formulas which do not contain* $>$ *is denoted by* $\mathcal{F}_0$.

Definition 7 (Degree of a formula) *The degree of a formula* ϕ *is denoted by* $deg(\phi)$ *and is defined as follows:*

- *For* $\phi \in \mathcal{F}_0$, $deg(\phi) = 0$*;*
- *For* $\phi \in \{\neg\psi, \mathrm{F}\psi, \mathrm{P}\psi, \Diamond\psi\}$, $deg(\phi) = deg(\psi)$*;*
- *For* $\phi = \psi \wedge \chi$, $deg(\phi) = max(deg(\psi), deg(\chi))$*;*
- *For* $\phi = \psi > \chi$, $deg(\phi) = max(deg(\psi), deg(\chi)) + 1$.

The symbol $\mathcal{F}_n$ *denotes the set of all formulas of degree up to – and including –* n.

B.1 Base step

- $S_0 \coloneqq \big\{s : \mathcal{F}_0 \times M/H \to M/H \mid \forall m/h \in M/H, \phi \in \mathcal{F}_0$
  - $s(\phi, m/h) \vDash_{BT} \phi$;
  - $s(\phi, m/h) \sim m/h$;
  - $\forall m''/h'' \in M((m'' \sim m) \wedge (h \cap h' \subset h \cap h'') \Rightarrow m''/h'' \nvDash_{BT} \phi),$ *where* h' *is the history in* $s(\phi, m/h)\big\}$.

The parameters for $\vDash_0$ are moment-history pairs from M and selection functions from S_0; the full definition of $\vDash_0$ is as follows, for $\phi \in \mathcal{F}_0$:

Definition 8 (0-truth)
$m/h, s \vDash_0 \phi$ *iff* $m/h \vDash_{BT} \phi$

B.2 Inductive step

In what follows, $s|_n$ is the restriction of the selection function s to formulas in $\mathcal{F}_n$.

Definition 9 (Class of all n-selection functions S_n) *An n-selection function is a function $s : \mathcal{F}_n \times M/H \mapsto M/H$ such that:*

1. $s|_{n-1} \in S_{n-1}$
2. *for any $\phi \in \mathcal{F}_n$ and $m/h \in M/H$:*
   - $s(\phi, m/h), s|_{n-1} \vDash_n \phi$;
   - $s(\phi, m/h) \sim m/h$;
   - $\forall_{m''/h'' \in M/H}((m'' \sim m) \wedge (h \cap h' \subset h \cap h'') \Rightarrow$ $m''/h'', s|_{n-1} \nvDash_n \phi)$ *where h' is the history in $s(\phi, m/h)\}$*

We can now proceed with the "meat" of the induction step. Assume that $\vDash_{n-1}$ and S_{n-1} are given and that $\phi \in \mathcal{F}_n$. The parameters for $\vDash_n$ are moment-history pairs and and selection functions from S_{n-1}.

Definition 10 (n-truth)

- *for $\phi \in \mathcal{F}_{n-1}$:*

  $m/h, s \vDash_n \phi$ *iff* $m/h, s \vDash_{n-1} \phi$;

- *for ϕ, $\psi \in \mathcal{F}_{n-1}$:*

  $m/h, s \vDash_n \phi > \psi$ *iff* $s(\phi, m/h), s \vDash_{n-1} \psi$;

- *for $\phi, \psi \in \mathcal{F}_n \setminus \mathcal{F}_{n-1}$:*

  - $m/h, s \vDash_n \neg\phi$ *iff* $m/h, s \nvDash_n \phi$
  - $m/h, s \vDash_n \psi \wedge \phi$ *iff* $m/h, s \vDash_n \psi$ *and* $m/h, s \vDash_n \phi$
  - $m/h, s \vDash_n \mathrm{F}\phi$ *iff* $\exists_{m'>m} m'/h, s \vDash_n \phi$
  - $m/h, s \vDash_n \mathrm{P}\phi$ *iff* $\exists_{m'<m} m'/h, s \vDash_n \phi$
  - $m/h, s \vDash_n \Diamond\phi$ *iff* $\exists s' \in S_{n-1} \exists h' \in H\ m/h', s' \vDash_n \phi$

B.3 Final definitions

We will now construct the "general" selection functions, that is, functions which are not defined for formulas up to some particular degree, but for arbitrary formulas. Each member of the set S^+ defined below is a "proto-selection function", that is, a set of selection functions containing exactly a single function of each degree such that for any two of those one is an extension of the other.

Definition 11 (Set S^+ of all proto-selection functions)
$S^+ := \{s^+ \subset \bigcup_{n \in \mathbb{N}} S_n |$

- $\forall_n \exists !_s$ *such that* $s \in S_n$ *and* $s \in s^+$
- $\forall_{s', s'', m}$ *if* $s', s'' \in s^+$, *and* $s' \in S_m$, *and* $s'' \in S_{m+1}$, *then* $s''|_m = s'\}$

The required selection functions are simply unions of proto-selection functions:

Definition 12 (Set S of all selection functions) $S := \{s | s = \bigcup s^+$, *for some* $s^+ \in S^+\}$

By relativization of truth to a choice of a selection function $s \in S$ we arrive at our ultimate definition of truth (we omit as obvious the parts concerning classical and temporal connectives):

Definition 13 (Truth)

- $m/h, s \vDash p$ *iff* $m/h \vDash_{BT} p$
- $\ldots$
- $m/h, s \vDash \phi > \psi$ *iff* $s(\phi, m/h), s \vDash \psi$
- $m/h, s \vDash \Diamond\phi$ *iff* $\exists_{s' \in S} \exists_{h' \in H}\ m/h', s' \vDash \phi$

C Reasoning with modalities and counterfactuals

Let us present some results of our explorations of the interplay of historical counterfactuals with historical necessity. To begin with, consider the pair of sentences:

- It is possible that if I had met a person in need yesterday, I would have helped.

- If I had met a person in need yesterday, it would have been possible that I would help.

These two seem to be synonymous; they seem to be true (or false) together and for the very same reason: our readiness (or lack thereof) to help.

Examples like these suggest that, as far as historical possibility is concerned, it makes no difference whether the modal operator takes a wide or a narrow scope with respect to the counterfactual; this would indicate the validity of the following rules of inference:

$$\text{(narrow-to-wide)} \ \frac{\phi > \Diamond\psi}{\Diamond(\phi > \psi)} \qquad \text{(wide-to-narrow)} \ \frac{\Diamond(\phi > \psi)}{\phi > \Diamond\psi}$$

If these rules were valid, it would open a new perspective on the debate regarding the scope ambiguity of might-counterfactuals. Prima facie, the sentence like "Had I flipped the coin, it might have landed heads" can be understood either as $\phi > \Diamond\psi$ or as $\Diamond(\phi > \psi)$. Given the validity of the rules, the scope ambiguity would make no difference, at least as long the historical modalities were concerned. With our semantic definitions, we can establish that in most cases, such two sentences are indeed equivalent, but in full generality, neither of the two inferences are valid. The following two examples show the invalidity of the two above rules under our semantics.

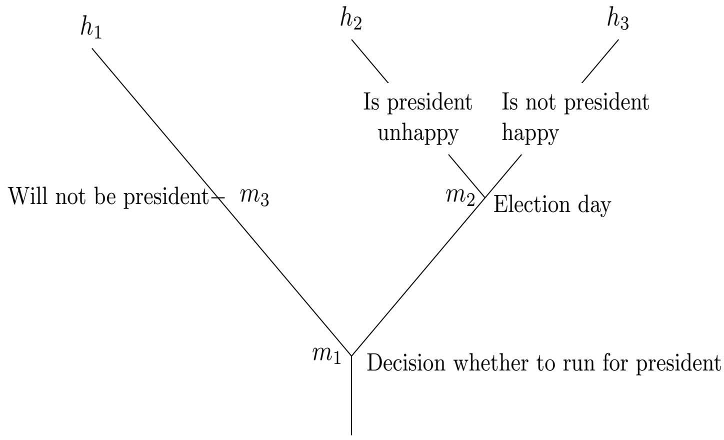

Figure 2: Potential election
1

The first case, illustrating the failure of the narrow-to-wide rule, uses an antecedent which is counterfactually contingent; that is, it is a sentence in the future tense which is not settled at the moment-history pair we "jump to" using the selection function. Consider Figure 2; the model obviously uses a lot of idealization but is just an illustration of a technical point. Suppose we had made the decision not to run for president; m_3 occurs on the date of the election in which, in h_1, we do not take part in the election. Consider the sentence, "If I were to become president, I might be happy" (F president $>$ $\Diamond$F happy).

For an arbitrary selection function s, we have that $s(\text{F president}, m_3/h_1) = m_2/h_2$, where the sentence "$\Diamond$F happy" is true (due to the existence of h_3, where we turn out to be not elected, but happy). However, keeping the same s, we see that at $m_3/h_1, s$ the sentence "$\Diamond$(F president $>$ F happy)" is false: there is no selection function which would make us jump to moment-history pair in which both "F president" and "F happy" hold. Thus, it is not possible that if I were to be elected, I would be unhappy. Therefore the narrow-to-wide principle fails in full generality; however, it holds when we restrict our attention to counterfactuals with antecedents speaking about the settled past (or, in general, antecedents ϕ such that $\phi > \Box\phi$).

The converse inference fails for an unrelated reason. Consider a model which illustrates the following scenario (Figure 3). There are two rigged coins in a box, a double-headed coin and a double-tailed coin. At moment m_0, I can draw and later toss one of the coins from the box. In history h_1, I draw a double-headed coin, in history h_3, I draw a double-tailed one and in history h_2, which is the actual one, I decide not to draw any coin at all. Let us take a moment m_2 in history h_2 and a selection function s such that $s(\text{toss}, m_2/h_2) = m_3/h_3$. Then we have that $m_2/h_2, s \vDash \Diamond(\text{toss} > \text{F(heads)})$, but $m_2/h_2, s \nvDash \text{toss} > \Diamond\text{F(heads)}$.

We have stated the rules in terms of $\Diamond$, but they can be equally well stated in terms of $\Box$ as the following two equivalences hold:

$$\frac{\phi > \Diamond\psi}{\Diamond(\phi > \psi)} \textit{ iff } \frac{\Box(\phi > \psi)}{\phi > \Box\psi} \quad \text{and} \quad \frac{\Diamond(\phi > \psi)}{\phi > \Diamond\psi} \textit{ iff } \frac{\phi > \Box\psi}{\Box(\phi > \psi)}$$

Since we cannot get the most general rules of inference we pursued, we can try the "second best". Consider the following strengthening of the so called "Edelberg inferences" advocated by Thomason and Gupta (1980):

$$\text{(Edelberg)} \ \frac{\Box(\phi > \psi)}{\phi > \Box(\phi \to \psi)} \qquad \text{(Weak Edelberg)} \ \frac{\phi > \Box\phi, \Box(\phi > \psi)}{\phi > \Box\psi}$$

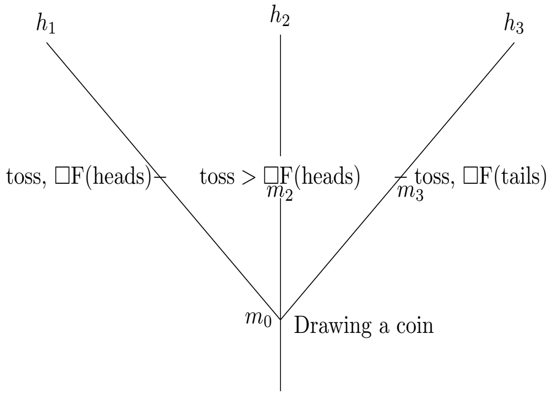

Figure 3: Two rigged coins
1

To validate these, the authors replaced the notion of history with a highly complex concept of a "future choice function", and then impose as many as 9 conditions on selection functions. Interestingly, the three simple requirements on selection functions we proposed above, together with our definition of possibility, are sufficient to validate both these inferences (we omit the formal details). We take it to be an advantage of our theory over the theory of Thomason and Gupta.

References

Belnap, N., Perloff, M., & Xu, M. (2001). *Facing the Future: Agents and Choices in Our Indeterministic World*. Oxford: Oxford University Press.

Kaplan, D. (1989). Afterthoughts. In J. Almong, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 565–614). Oxford: Oxford University Press.

Leitgeb, H. (2012). A Probabilistic Semantics for Counterfactuals. Part A. *The Review of Symbolic Logic*, *5*, 26–84.

Lewis, D. (1986). *Philosophical Papers* (Vol. II). Oxford: Oxford University Press.

MacFarlane, J. (2003). Future Contingents and relative Truth. *The Philosophical Quarterly*, *53*(212), 321–336.

MacFarlane, J. (2014). *Assessment Sensitibity: Relative Truth and Its Aplications*. Oxford: Clarendon Press.

Placek, T., & Müller, T. (2007). Counterfactuals and Historical Possibility. *Synthese*, *154*, 173–197.

Pollock, J. (1976). *Subjunctive Reasoning* (Vol. 28) (No. 113). Reidel.

Stalnaker, R. (1968). A Theory of Conditionals. In N. Rescher (Ed.), *Studies in Logical Theory* (pp. 23–42). Blackwell.

Stalnaker, R. (1981). A defense of conditional excluded middle. In R. S. W. Harper & G. Pearce (Eds.), *Ifs* (pp. 87–104). Dordrecht.

Thomason, R. H. (1970). Indeterminist Time and Truth-value Gaps. *Theoria*, *36*, 264–281.

Thomason, R. H., & Gupta, A. (1980). A Theory of Conditionals in the Context of Branching Time. *The Philosophical Review*, *89*(1), 65–90.

Jacek Wawer
Jagiellonian University, Kraków
Poland
E-mail: jacek.wawer@fulbrightmail.org

Leszek Wroński
Jagiellonian University, Kraków
Poland
E-mail: leszek.wronski@uj.edu.pl