

# An interactive demonstration of counterfactual truth conditions

Bachelor Thesis

*Andreas Paul Bruno Lönne*  
*loenne@campus.tu-berlin.de*

Technische Universität Berlin  
discourse Degree program: Bachelor Informatik / Computer Science

## Abstract

In this thesis, I address the scarcity of online resources, showcasing counterfactual truth conditions in an intuitive and digestible manner. To this end I (i) formulate a semantic satisfiability game for counterfactual sentences; (ii) prove its correctness; (iii) prove that it always halts after a finite number of moves; (iv) develop a browserbased web-application, that makes the semantic game of counterfactuals playable.

## 1 Counterfactual logic

This section is concerned with defining counterfactual logic. First it provides a brief definition of well-formed counterfactual formulas, then introduces the structure they are evaluated on and concludes by defining the truth conditions of counterfactual logic.

### 1.1 Counterfactual formulas

We call  $\Phi = \{\varphi, \psi, \dots\}$  the set of all well-formed counterfactual formulas.

**Definition 1** (Well-formed counterfactual formula). Given an infinite set of atomic formula symbols  $Atoms = \{x, y, \dots\}$  and an alphabet  $A = \{\perp, \top, \neg, \vee, \wedge, \Diamond, \Box, \Box \rightarrow, \Diamond \rightarrow\} \cup Atoms$ , the structure of every well-formed counterfactual formula is expressed through the following Backus-Naur form.

$$\varphi, \psi ::= \perp \mid \top \mid x \mid \neg\varphi \mid \varphi \vee \psi \mid \varphi \wedge \psi \mid \Diamond\varphi \mid \Box\varphi \mid \varphi \Box \rightarrow \psi \mid \varphi \Diamond \rightarrow \psi \quad (1)$$

### 1.2 Counterfactual kripke structure

In order to evaluate the non-truth-functional connectives of counterfactual logic, we define a variation of the Kripke structure in [?]. To evaluate Lewis' counterfactual truth conditions [Lew73], we also introduce a notion of similarity between possible worlds to the accessibility relation.

**Definition 2** (counterfactual kripke structure). A *counterfactual kripke structure* is an ordered triple  $(W, \rightsquigarrow, F)$ , where  $W = \{w, v, \dots\}$  is the set of all possible worlds,  $\rightsquigarrow: W \times \mathbb{R} \times W$  is the similarity relation, and  $F: W \rightarrow 2^{Atoms}$  is an assignment of each world, to a set of atomic formulas.

Let us explain further the members of our counterfactual kripke structure.

The set of all possible worlds  $W$  represents all complete and self-consistent ways reality could have been.  $F$  in turn, describes the state of affairs at each world. It assigns to each world a set of atomic propositions, that are the case at it. The similarity relations purpose is twofold. It serves as an accessibility relation, restricting accessibility between worlds and also carries information about comparative similarity between worlds. In particular this means for any two worlds  $w$  and  $v$ , that iff  $w \rightsquigarrow^r v$ , then  $v$  is called accessible from  $w$ .  $w \rightsquigarrow^r v$  also means, that  $r$  is a measure of similarity between  $w$  and  $v$ , from the standpoint of  $w$ . We can suppose another world  $v'$ , with  $w \rightsquigarrow^{r'} v'$  and compare  $r$  and  $r'$ . If  $r < r'$ , we can conclude, that  $v$  is more similar to  $w$  than  $v'$ , from the standpoint of  $w$ . Additionally we note, that for our purposes  $\rightsquigarrow$  will always contain a tuple  $(w, 0, w)$  for each world  $w \in W$ . This is because we assume analogously to Kripke, that worlds are self-similar and are accessible from themselves. Finally we introduce  $W_w = \{w' \mid w \rightsquigarrow w'\}$  as a shorthand for the set of all accessible worlds, from a world  $w$ .

### 1.3 Truth conditions of counterfactual logic

Given a counterfactual kripke structure  $S = (W, \rightsquigarrow, F)$  and a world  $w$ ,

$S, w \not\models \perp$ .

$S, w \models \top$ .

$S, w \models x$  iff  $x \in F(w)$ .

$S, w \models \neg\varphi$  iff  $S, w \not\models \varphi$ .

$S, w \models \varphi \vee \psi$  iff  $(S, w \models \varphi \text{ or } S, w \models \psi)$ .

$S, w \models \varphi \wedge \psi$  iff  $(S, w \models \varphi \text{ and } S, w \models \psi)$ .

$S, w \models \Diamond\varphi$  iff a world  $w' \in W_w$  exists, such that  $S, w' \models \varphi$ .

$S, w \models \Box\varphi$  iff for every world  $w' \in W_w$ , it is true that  $S, w' \models \varphi$ .

$S, w \models \varphi \Box\rightarrow \psi$ , iff either

- (1) no world  $w' \in W_w$  exists, such that  $S, w' \models \varphi$ , or
- (2) a world  $w'$  and an  $r$  exist, such that  $w \rightsquigarrow^r w'$  and  $S, w' \models \varphi$  and for each world  $w^*$ , for which an  $r^*$  exists, such that  $r^* \leq r$  and  $w \rightsquigarrow^{r^*} w^*$ , it is true that  $S, w^* \models \neg\varphi \vee \psi$ .

$S, w \models \varphi \Diamond\rightarrow \psi$ , iff both

- (1) a world  $w' \in W_w$  exists, such that  $S, w' \models \varphi$ , and
- (2) for every world  $w''$ , for which an  $r''$  exists, such that  $w \rightsquigarrow^{r''} w''$  and  $S, w'' \models \varphi \wedge \neg\psi$  are true, a world  $w^*$  and an  $r^*$  exist, such that  $r^* \leq r''$  and  $w \rightsquigarrow^{r^*} w^*$  and  $S, w^* \models \varphi \wedge \psi$ .

## 2 The semantic game of counterfactuals

This section offers an overview and introduction to the semantic game of counterfactuals. The semantic game of counterfactuals is a sequential two-player reachability game, wherein

a *defender*  $d$  tries to prove a counterfactual formula and an *attacker*  $a$  tries to disprove it. The defender begins the game as the *active player*, with a counterfactual formula at a possible world. The active player is a role assigned to the player that can currently make moves and whose typical objective it is to prove the current formula through play. As the active player makes moves, the counterfactual formula is resolved progressively until a player resolves a  $\top$  and wins; or cannot make any move and loses. Throughout the game, moves may also change the active player and current world.

## 2.1 Semantic game tuple

The game is akin to a discussion where players make arguments and refute them. Weak arguments do not hold up to scrutiny, but strong ones do. The players' adversarial interplay allows the discovery and dismissal of arguments, such that the reason for truth or falsity of a claim is revealed. Of course, this is only possible when the players make the best possible moves. We shall call such play optimal play. We describe the game the following way.

**Definition 3** (Semantic game of counterfactuals). The semantic game of counterfactuals is an ordered tuple  $(I, E, L)$ , where  $I \in G$  is the initial game state;  $E = \{(\top, w)_p \mid w \in W, p \in \{a, d\}\}$  is the set of winning game states;  $L$  is a transition system for game states.

$G$  is the set of all game states. We will describe the structure of its elements in the upcoming subsections. The initial game state  $I$  varies from game to game. It contains the formula that ought to be proven and the world it ought to be proven at. It is always indexed by the initial active player, that is the defender. The set of winning game states contain all game states in which a player wins. In each winning game state it is the currently active player that wins and the respective other player that loses. This set includes all game states where the formula is resolved to a  $\top$ -symbol. It is also possible for a player to win without reaching a winning game state. This is the case when the active player is unable to make a move and in other words stuck. When this happens, the active player loses and the active player's opponent i.e. the inactive player wins. Next we will define the transition system  $L$ .

## 2.2 The transition system

The moves that can be played in a given game state are determined by a transition system  $L$ .

**Definition 4** (Transition system). A transition system is a tuple  $(G, R)$ , where  $G$  is the set of all game states and  $R : G \times G$  is a transition relation.

It contains the set of all game states  $G$  and the set  $R$  of transitions between game states which we call moves. In later sections we will define two such sets of moves. One general formulation  $\rightarrow$  and one simplified formulation  $\rightarrow_l$  employing the limit assumption. One may take either as  $R$ . In order to arrive at our definition for  $\rightarrow$  and  $\rightarrow_l$ , let us first define game states and transitions.

## 2.3 Game state

Game states form the backbone of the semantic game of counterfactuals by describing its game positions. Each game state describes the game at a discrete point in time in-between moves. We call the set of all game states  $G$ .

**Definition 5** (Game state). A game state is a tuple that can take any one of the forms

- (1)  $(\varphi, w)_p$
- (2)  $(\varphi, w, e)_p$
- (3)  $(\varphi, w, w', r)_p$

where  $\varphi \in \Phi$ ;  $w, w' \in W$ ;  $r \in \mathbb{R}$  and  $p, e \in \{a, d\}$ .

Of these types of game states (1) is the most common. It carries a counterfactual formula, the current world and is indexed with the active player. It is employed to resolve atomic formulas  $x, y, \dots$  and the propositional symbols  $\top, \neg, \vee, \diamond$ ; while also serving as an initial and final state type for other resolutions. (2) is a type of game state that represents a player temporarily becoming active and choosing a state transition.  $e$  holds the previous active player, which will resume being active after the next state transition. Lastly game state type (3) is an extension of (1), describing a game state where a sphere of accessibility has been chosen.  $w'$  is a *sphere-delimiting world*, lying exactly on the surface of said sphere of accessibility with radius  $r$ . We will look at this in further detail in the next subsection.

## 2.4 Moves

When the active player transforms the current game state into another and progresses the game, we call that a move. Formally we represent this through a tuple of game states.

**Definition 6** (Move). A move  $m$  is a tuple  $m \in G \times G$ .

The first element of the tuple is the initial game state, while the second one is the resulting game state. We define the set of moves  $\rightarrow$  specific to the semantic game of counterfactuals.

**Definition 7** (Moves).  $\rightarrow$  is a set of moves, such that it contains exactly those tuples that fit any of the forms described by the following 19 types of moves. For each of those types of moves variables are quantified separately. Additionally, square brackets above the relation symbol are used to annotate restrictions imposed upon the quantification of those variables. In case of (1) for example  $x$  is restricted to range across atomic formulas contained in  $F(w)$ . Furthermore, it is always true that  $\varphi, \psi \in \Phi$  are counterfactual formulas;  $x \in Atoms$  is an atomic formula;  $w, w', w^* \in W$  are worlds;  $p, e \in \{a, d\}$  are players;  $np \in \{a, d\} \setminus \{p\}$  is the opponent of the player  $p$ ; and  $r, r^* \in \mathbb{R}$ .

$$\begin{aligned}
(x, w)_p &\xrightarrow{[x \in F(w)]} (\top, w)_p & (2) \\
(\neg\varphi, w)_p &\rightarrow (\varphi, w)_{np} & (3) \\
(\varphi \vee \psi, w)_p &\rightarrow (\varphi, w)_p & (4) \\
(\varphi \vee \psi, w)_p &\rightarrow (\psi, w)_p & (5) \\
(\varphi \wedge \psi, w)_p &\rightarrow (\varphi \wedge \psi, w, p)_{np} & (6) \\
(\varphi \wedge \psi, w, e)_p &\rightarrow (\varphi, w)_e & (7) \\
(\varphi \wedge \psi, w, e)_p &\rightarrow (\psi, w)_e & (8) \\
(\Diamond\varphi, w)_p &\xrightarrow{[w \prec w']} (\varphi, w')_p & (9) \\
(\Box\varphi, w)_p &\rightarrow (\Box\varphi, w, p)_{np} & (10) \\
(\Box\varphi, w, e)_p &\xrightarrow{[w \prec^r w']} (\varphi, w')_e & (11) \\
(\varphi \Box\rightarrow \psi, w)_p &\rightarrow (\Box\neg\varphi, w)_p & (12) \\
(\varphi \Box\rightarrow \psi, w)_p &\xrightarrow{[w \prec^r w']} (\varphi \Box\rightarrow \psi, w, w', r)_{np} & (13) \\
(\varphi \Box\rightarrow \psi, w, w', r)_p &\rightarrow (\varphi, w')_{np} & (14) \\
(\varphi \Box\rightarrow \psi, w, w', r)_p &\xrightarrow{[w \prec^r w', r^* \leq r]} (\neg\varphi \vee \psi, w^*)_{np} & (15) \\
(\varphi \Diamond\rightarrow \psi, w)_p &\rightarrow (\varphi \Diamond\rightarrow \psi, w, p)_{np} & (16) \\
(\varphi \Diamond\rightarrow \psi, w, e)_p &\rightarrow (\Box\neg\varphi, w)_p & (17) \\
(\varphi \Diamond\rightarrow \psi, w, e)_p &\xrightarrow{[w \prec^r w']} (\varphi \Diamond\rightarrow \psi, w, w', r)_e & (18) \\
(\varphi \Diamond\rightarrow \psi, w, w', r)_p &\rightarrow (\varphi, w')_{np} & (19) \\
(\varphi \Diamond\rightarrow \psi, w, w', r)_p &\xrightarrow{[w \prec^r w', r^* \leq r]} (\varphi \wedge \psi, w^*)_p & (20)
\end{aligned}$$

Fig. 1: Moves of the semantic game of counterfactuals

Each symbol of the counterfactual logic is resolved through one or more resolution steps. Take (9) and (10) for the resolution of the necessity operator as an example. The operator is resolved in two steps by first making the move (9) and then (10). Each such sequence of resolution steps both begins and ends with a game state of the form  $(\varphi, w)_p$ . The meaning of this type of game state is that the active player  $p$  attempts to show the formula  $\varphi$  at the world  $w$  through subsequent play. Accordingly, the game's moves are defined in such a way that the player  $p$  wins the game with optimal play, iff  $\varphi$  is true at  $w$ . The formula  $\varphi$  is resolved top-down, by continually resolving the weakest-binding operator or atomic formula, until no further resolution is possible. Since the truth conditions of the operators of counterfactual logic depend on the truth of the subformulas they bind, a game to determine the truth of the relevant subformula—or in some cases a different formula—is played

after each resolution of a counterfactual operator. For this reason all resolution sequences of all operators and atomic formulas end with the type of game state  $(\varphi, w)_p$  that describes the initial state of a new game pertaining to the truth of the formula  $\varphi$ . This allows us to consider the resolution sequences for each operator separately. Given the invariant that the player  $p$  wins the game with optimal play from the game state  $(\varphi, w)_p$ , iff  $\varphi$  is true at  $w$  for each subformula  $\varphi$ . Because players frequently switch the role of active player throughout resolution sequences, it will become cumbersome and confusing to describe players by calling them the active or inactive player. We will thus call the player that is active at the beginning of a resolution sequence the proving player and their opponent the disproving player.

(1) Atomic formulas which are true at the current world are resolved to a  $\top$ -symbol. The proving player subsequently reaches the  $\top$ -symbol and wins the game. In any game state where the current formula is an atomic formula that is not true at the current world the proving player loses since no move is possible and they are stuck.

(2) The negation is resolved by switching the active player. The reason for this is that the truth of the formula  $\neg\varphi$  can be disproven by proving the formula  $\varphi$ . Thus this move switches the active player and makes the disproving player prove the subformula  $\varphi$ . If they are able to do so, they win and the proving player loses. Otherwise they lose and the proving player wins the game.

(3), (4) The disjunction is resolved through the proving player's choice of which subformula to prove i.e. the choice to make the move (3) or (4). If one of the subformulas is true, then the proving player can choose it, prove it through further play and win. In any other case the proving player loses with further optimal play.

(5), (6), (7) Conjunctions are resolved by making the disproving player temporarily become the active player and resolve the conjunction as if it were a disjunction. In other words, the disproving player becomes active and chooses which subformula the proving player has to prove through play, before becoming inactive again. With optimal play the disproving player chooses the harder subformula for their opponent to prove. If one of the subformulas cannot be proven through play, the disproving player chooses it and wins subsequently with optimal play. Conversely if such a subformula does not exist.

(8) The possibility operator is resolved by having any accessible world become the new current world. The proving player gets to choose that world and has to subsequently prove through play that the subformula  $\varphi$  is true at it. If there exists any accessible  $\varphi$ -world, then the proving player can choose that world, show  $\varphi$  there through play and win the game. If no such world exists, the proving player has to choose a world  $\varphi$  cannot be proven at and loses subsequently with optimal play.

(9), (10) The resolution of the necessity operator relates to the resolution of the possibility operator in a similar way as the resolution of conjunctions relates to the resolution of disjunctions. Both employ a paradigm where the relationship of duality with another operator is used by switching the active player and resolving it like their respective counterpart operator. First the disproving player becomes temporarily active (9), resolves necessity as if it were possibility (10), becomes inactive again and leaves the proving player to prove the subformula  $\varphi$  at the chosen world. This is because necessity is true, when it's supposition is true at all accessible worlds. Thus the disproving player makes the worst choice in the proving player's stead. If there exists any accessible world at which the subformula  $\varphi$  is not true, it is sure to be chosen with optimal play. The proving player then cannot prove  $\varphi$  there through play and loses the game. If  $\varphi$  is true at every accessible world however, then no world can be chosen, such that the proving player cannot prove  $\varphi$  there through play. The proving player wins the game with optimal play in that case.

(11), (12), (13), (14) Sticking closely to the structure of Lewis' definition the counterfactual would's truth conditions, the proving player first decides which truth condition—vacuous or non-vacuous truth—to claim truth of and force their opponent to attempt to disprove. Move type (11) is the choice of vacuous truth and (12) non-vacuous truth. When the proving player claims vacuous truth, they are required to prove the formula  $\Box \neg \varphi$ . This means that every accessible world has to not be a  $\varphi$ -world. If an accessible  $\varphi$ -world exists, the disproving player can choose it as per  $\Box$ , become the active player as per  $\neg$ , then prove  $\varphi$  there and win the game. If no such world exists, the disproving player will fail to prove  $\varphi$  with optimal play and lose.

When deciding to claim that the counterfactual would in question is non-vacuously true, the initial active player also chooses a sphere of accessibility around the current world that is meant to prove non-vacuous truth. This sphere has to contain at least one  $\varphi$ -world and throughout each world within it  $\varphi \rightarrow \psi$  has to hold. For the semantic game we adopt a slightly different but unequivocally equivalent condition. We give a sphere around a world  $i$  through a world that is a most dissimilar world to  $i$  still contained within the sphere. We call such a world a sphere-delimiting or delimiting world. We say that the chosen sphere's delimiting world has to be a  $\varphi$ -world and that  $\varphi \rightarrow \psi$  has to hold throughout each world of that sphere. We know that whenever there is a sphere containing a  $\varphi$ -world throughout which  $\varphi \rightarrow \psi$  holds, then there also is a subset of that sphere that is a sphere with a delimiting  $\varphi$ -world throughout which  $\varphi \rightarrow \psi$  holds. We can simply take the  $\varphi$ -world of the former sphere as the delimiting world for the latter. And whenever there is a sphere with a delimiting  $\varphi$ -world throughout which  $\varphi \rightarrow \psi$  holds, then that sphere is also a sphere containing a  $\varphi$ -world throughout which  $\varphi \rightarrow \psi$  holds. Now that we have introduced the semantic game's notion of the counterfactual would's non-vacuous truth, let us explain the corresponding moves (13) and (14). After the proving player's choice to claim that the counterfactual is non-vacuously true, the disproving player chooses in which way to attempt to disprove that claim. They can claim that the chosen sphere-delimiting world is not a  $\varphi$ -world (13) and force the initial active player to subsequently prove it through play. Or they can assert that  $\varphi \rightarrow \psi$  does not hold throughout the chosen sphere (14) and choose a world serving as a counterexample. The proving player then has to prove that  $\varphi \rightarrow \psi$  holds at that world by proving through play that  $\neg \varphi \vee \psi$  holds at it.

(15), (16), (17), (18), (19) In a similar vain to the conjunction and necessity, the counterfactual might's resolution hinges on it's interdefinability with the counterfactual would. Since the interdefinition of the counterfactual would and might contains two negations however, it's resolution sequence does not show such a clear correspondence of the counterfactual would's. The counterfactual might is resolved by switching the active player temporarily (15) until the choice between (16) and (17) has been made and follows the same structure as the resolution sequence of the counterfactual would otherwise.

## 2.5 Limit Assumption

Additionally, Lewis states the limit assumption which allows a simplification of the semantic game of counterfactuals. We will call a sphere that contains a  $\varphi$ -world a  $\varphi$ -permitting sphere.

**Definition 8** (Limit assumption). For every world  $i$  and formula  $\varphi$  for which a  $\varphi$ -permitting sphere around  $i$  exists, there is a smallest  $\varphi$ -permitting sphere around  $i$ .

In other words, a nonempty set of  $\varphi$ -permitting spheres around  $i$  has a smallest member. Without a doubt this is true when there are only finitely many spheres around  $i$ . We can

reassure us of that fact the following way. Given any sphere of a nonempty set of  $\varphi$ -permitting spheres  $\$i$  around a world  $i$ , we can check for each sphere in  $\$i$  whether it is smaller than that sphere. If that is the case for any sphere, we can take this sphere as our next sphere and repeat the same process for it. If we do not find a smaller sphere, then we found a smallest sphere. And since we assumed  $\$i$  to be finite, we will invariably run out of smaller spheres to find. If we take  $\$i$  to be an infinite set however, we may find an infinite descending sequence of smaller and smaller spheres without end. In that case, we cannot find a smallest sphere, since every sphere has a sphere that is smaller than it.

Since the spheres around  $i$  are nested, one of any two spheres contains the other. It follows that the smallest sphere of a set of spheres  $\$i$ , if it exists, is the intersection of all spheres of  $\$i$ . This means that it is contained in all other spheres in  $\$i$ . This circumstance proves quite useful to reduce the semantic game's complexity. This is because showing that a formula  $\varphi$  does not hold throughout a smallest sphere around a world  $i$  also shows that it does not hold throughout any larger sphere and by extension any sphere in  $\$i$ . We will apply this notion to simplify the semantic game's resolution of counterfactual operators and thereby provide an alternative formulation of it by employing the limit assumption.

## 2.6 Alternative moves

**Definition 9** (Simplified moves).  $\rightarrow_i$  is a set of moves, such that it contains exactly those tuples that fit any of the forms described by the following 17 types of moves. Variables are quantified the same way as in definition 7.



$$\begin{aligned}
(x, w)_p &\xrightarrow{[x \in F(w)]} (\top, w)_p & (21) \\
(\neg\varphi, w)_p &\rightarrow (\varphi, w)_{np} & (22) \\
(\varphi \vee \psi, w)_p &\rightarrow (\varphi, w)_p & (23) \\
(\varphi \vee \psi, w)_p &\rightarrow (\psi, w)_p & (24) \\
(\varphi \wedge \psi, w)_p &\rightarrow (\varphi \wedge \psi, w, p)_{np} & (25) \\
(\varphi \wedge \psi, w, e)_p &\rightarrow (\varphi, w)_e & (26) \\
(\varphi \wedge \psi, w, e)_p &\rightarrow (\psi, w)_e & (27) \\
(\Diamond\varphi, w)_p &\xrightarrow{[w \prec w']} (\varphi, w')_p & (28) \\
(\Box\varphi, w)_p &\rightarrow (\Box\varphi, w, p)_{np} & (29) \\
(\Box\varphi, w, e)_p &\xrightarrow{[w \prec w']} (\varphi, w')_e & (30) \\
(\varphi \Diamond\rightarrow \psi, w)_p &\xrightarrow{[w \prec w']} (\varphi \Diamond\rightarrow \psi, w, w', r)_{np} & (31) \\
(\varphi \Diamond\rightarrow \psi, w, w', r)_p &\rightarrow (\varphi \wedge \psi, w')_{np} & (32) \\
(\varphi \Diamond\rightarrow \psi, w, w', r)_p &\xrightarrow{[w \prec w^*, r^* < r]} (\neg\varphi, w^*)_{np} & (33) \\
(\varphi \Box\rightarrow \psi, w)_p &\rightarrow (\varphi \Box\rightarrow \psi, w, p)_{np} & (34) \\
(\varphi \Box\rightarrow \psi, w, e)_p &\xrightarrow{[w \prec w']} (\varphi \Box\rightarrow \psi, w, w', r)_e & (35) \\
(\varphi \Box\rightarrow \psi, w, w', r)_p &\rightarrow (\neg\varphi \vee \psi, w')_p & (36) \\
(\varphi \Box\rightarrow \psi, w, w', r)_p &\xrightarrow{[w \prec w^*, r^* < r]} (\varphi, w^*)_p & (37)
\end{aligned}$$

Fig. 2: Moves employing the limit assumption

Moves (1)-(10) are the same as in the formulation in figure 1.

(11), (12), (13) With respect to the limit assumption, we can formulate a simpler resolution of the counterfactual might. It can be proven by showing that the smallest  $\varphi$ -permitting sphere around the current world  $w$  does contain a  $(\varphi \wedge \psi)$ -world. In that case every larger sphere around  $w$  also contains a  $(\varphi \wedge \psi)$ -world; and every smaller sphere around  $w$  does not contain a  $\varphi$ -world. This corresponds to the requirement (2) of Lewis' truth conditions of the counterfactual might (see figure ??). Furthermore, it is necessary to show that a  $\varphi$ -permitting sphere around  $w$  exists. Hence the counterfactual might's resolution sequence begins with the proving player's choice of a sphere of accessibility (11). When the proving player cannot choose a sphere, because no other world is accessible, they lose. In that case, no  $\varphi$ -permitting sphere around  $w$  exists. In the case that the proving player is able to choose a sphere-delimiting world and thus a sphere, the chosen world ought to be a  $(\varphi \wedge \psi)$ -world. This is because we combine two truth requirements that the sphere needs to contain a  $\varphi$ -world; and that the sphere needs to contain a  $(\varphi \wedge \psi)$ -world. If the latter is true, the former is

true as well. Thus we can omit the less stringent requirement. Analogously to the previous formulation of the counterfactual would's resolution, we assert without loss of generality that the chosen sphere be  $(\varphi \wedge \psi)$ -delimited instead of merely containing a  $(\varphi \wedge \psi)$ -world. If the proving player did not choose a  $(\varphi \wedge \psi)$ -world, then the disproving player is able to refute that by forcing the proving player to prove  $(\varphi \wedge \psi)$  at the chosen world (12). If the delimiting world is either not a  $\varphi$ -world or not  $(\varphi \wedge \psi)$ -world, then the proving player loses subsequently with optimal play. And if there exists a smaller  $\varphi$ -permitting sphere than the sphere chosen by the proving player, then that players opponent can choose a closer world and force the proving player to attempt to show through play that the chosen world is not a  $\varphi$ -world (13). If it is a  $\varphi$ -world, then the proving player loses with optimal play and wins otherwise.

(14), (15), (16), (17) The counterfactual would is resolved by switching the active player before and after the choice of a sphere of accessibility and resolving it similarly to a counterfactual might. First, the active player is switched (14). Then the disproving player chooses a sphere of accessibility (15). And finally the proving player becomes active again and chooses in which way to disprove the disproving players chosen sphere. The proving player may either prove that  $\varphi \rightarrow \psi$  holds at the chosen sphere-delimiting world by proving that  $\neg\varphi \vee \psi$  holds there (16) or they may prove that there is a smaller  $\varphi$ -permitting sphere by proving that there is a more similar  $\varphi$ -world (17).

## References

- [Kri63] Saul A. Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16:83–94, 1963.
- [Lew73] David K. Lewis. *Counterfactuals*. Basil Blackwell, Oxford, 1973.