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 Introduction

In this thesis i make the attempt to create an application that is able to convey Lewis' counterfactual truthconditions by way of a semantic two-player game and inspire players to learn more about counterfactuals.
Background and literature
To this end i defined a semantic game of counterfactuals and implemented a browserbased demonstration game.
The document is laid out as follows:
First i will explain counterfactuals.
First i will begin by introducing the counterfactual logic im basing the semantic game of counterfactuals on. Then i will give game-theoretical definitions and formulate two versions of the semantic game. After that ...

2 Counterfactuals

Counterfactuals are statements about what might or would have been the case, if things took place differently than they did. One may think to themselves "If I had not forgotten about my appointment, I would have been punctual". Or one may wonder "If Alexander the Great had not died at the age of 32 and attacked europe, would the Romans have defeated him?". Such *counterfactual thought*—that is the thought of alternate outcomes—is essential for reasoning, deduction and cognitive function. [**?**] Due to its abundance in human thought, the ability to imagine alternate realities seems trivial to most. But making rigorous statements or claims about them is difficult. This is because communicating a

complete and consistent account of the state of affairs of an alternate reality—similar in complexity to ours— is difficult, if not impossible. One may attempt to circumvent this issue by giving the state of affairs of an alternate reality as a deviation from the state of affairs of reality. But consider this. Take our previous example about Alexander the Great and assume that the imagined alternate reality is identical to our reality, except that Alexander the Great did not die at the age of 32 and attacked europe. Now suppose we know about reality, that Alexanders troops remained outside of europe. Then this should also be the case in the alternate reality we attempted to describe. If we are to assume that an army cannot be in two places at the same time and Alexander could not have attacked europe without his army, then Alexanders troops could not have remained outside of europe and attacked europe at the same time. We find, that simply deviating from our own reality in a few concrete ways may produce internally inconsistent alternate realities, which cannot be alternate realities, because they are not ways the world could have been.
To avoid this problem we forgo describing alternate realities alltogether. We call a complete and consistent way the world is or could have been a possible world, and agree that we refer to a possible world, most similar to ours, where our stipulation is true. So our example is to be read as "In a world most similar to our own, where Alexander the Great did not die at the age of 32 and attacked europe, the romans would have defeated him". In this way—although we may not know the state of affairs of a possible world—we are able to assign definite truth values to our counterfactual sentence for any possible state of affairs at that world. However we need to note, that this approach assumes the notion of comparative similarity between possible worlds. Which means that given any 3 worlds w, v_1, v_2, with respect to their overall similarity, either

- v_1 is more similar to w, than v_2,
- v_2 is more similar to w, than v_1,
- or v_1 and v_2 are equally similar to w.

While the notion of an aggregate overall similarity between possible worlds appears justified at first glance, it has been subject of contention. [**?**]

2.1 Lewis' counterfactual operators

With these introductory thoughts out of the way, let us talk in greater detail about the counterfactual operators themselves. Lewis introduces the counterfactual would $\Box\!\!\rightarrow$ and counterfactual might $\Diamond\!\!\rightarrow$ operators as binary modal operators. [**?**] When we write $\varphi \Box\!\!\rightarrow \psi$, we call φ the antecedent and ψ the consequent. We may informally rewrite one of our former examples as "I did not forget about my appointment $\Box\!\!\rightarrow$ I was punctual" and read it the following way.
Read $\varphi \Box\!\!\rightarrow \psi$ as "If it were the case that φ, then it would be the case that ψ", and
read $\varphi \Diamond\!\!\rightarrow \psi$ as "If it were the case that φ, then it might be the case that ψ".
Lewis defines the truth conditions of his operators with respect to a system of spheres, that is defined as follows.

Let $ be an assignment to each possible world i of a set $\$_i$ of sets of possible worlds. Then $ is called a (centered) system of spheres, and the members of each $\$_i$ are called spheres around i, if and only if, for each world i, the following conditions hold.

(C) $\$_i$ is centered on i; that is, the set $\{i\}$ having i as its only member belongs to $\$_i$.

(1) $\$_i$ is nested; that is, whenever S and T belong to $\$_i$, either S is included in T or T is included in S.

(2) $\$_i$ is closed under unions; that is, whenever $\mathcal{S}$ is a subset of $\$_i$ and $\bigcup \mathcal{S}$ is the set of all worlds j such that j belongs to some member of $\mathcal{S}$, $\bigcup \mathcal{S}$ belongs to $\$_i$.

(3) $\$_i$ is closed under (nonempty) intersections; that is, whenever $\mathcal{S}$ is a nonempty subset of $\$_i$ and $\bigcap \mathcal{S}$ is the set of all worlds j such that j belongs to every member of $\mathcal{S}$, $\bigcap \mathcal{S}$ belongs to $\$_i$.

Fig. 1: Lewis' (centered) system of spheres

And his counterfactual truth conditions are:

$\varphi \diamond\!\!\rightarrow \psi$ is true at a world i (according to a system of spheres $) if and only if both

(1) some φ-world belongs to some sphere S in $\$_i$, and

(2) every sphere S in $\$_i$ that contains at least one φ-world contains at least one world where $\varphi \wedge \psi$ holds.

Fig. 2: Lewis' counterfactual might truth conditions

$\varphi \mathrel{\Box\!\!\rightarrow} \psi$ is true at a world i (according to a system of spheres $) if and only if either

(1) no φ-world belongs to any sphere S in $\$_i$, or

(2) some sphere S in $\$_i$ does contain at least one φ-world, and $\varphi \rightarrow \psi$ holds at every world in S.

Fig. 3: Lewis' counterfactual would truth conditions

Where a world, where φ holds is called a φ-world.