I
Counterfactuals

– If Oswald did not kill Kennedy, then someone else did.

– If Oswald had not killed Kennedy, then someone else would have.[1]

The phrases above are respectively instances of the indicative and the subjunctive conditionals. The indicative conditional is associated to the material implication, whereas the subjunctive construction of the language is traditionally studied by the philosophy as the counterfactual conditional[1, 3] or the counterfactual for short.

Conditional propositions involve two components, the antecedent and the consequent. Counterfactual conditionals differ from material implication in a subtle way. The truth of a material implication is based on the actual state-of-affairs. From the knowledge that Kennedy was killed, we can accept the truth of the phrase. On the other hand, a counterfactual conditional should take into account the truth of the antecedent, even if it is not the case. The truth of the antecedent is mandatory in this analysis.

Some approaches to counterfactuals entail belief revision, particularly those based on *Ramsey* test evaluation [11]. In this analysis, the truth value of a counterfactual is considered within a minimal change generated by admitting the antecedent true[3].

A possible way to circumvent belief revision mechanisms is to consider alternative (possible) state-of-affairs, considered here as worlds, and, based on some accessibility notion, choose the closest one among the worlds that satisfy the antecedent. If the consequent is true at this considered world, then the counterfactual is also true[1].

Both conditionals have false antecedents and false consequents in the current state-of-affairs. However, the second conditional is clearly false, since we found no reason to accept that, in the closest worlds in which Kennedy is not killed by Oswald, Kennedy is killed by someone else.

We choose the approach of Lewis[1] in our attempt to formalize an inference system for counterfactuals because his accessibility relation leaves out the discussion for a general definition of similitude among worlds, which is considered as given in his analysis.

It also opened the possibility for a contribution in the other way. If we found some general properties in his accessibility relation, considering the evaluation of the formulas in the counterfactual reasoning, we could sketch some details of the concepts of similitude.

I.1 Lewis analysis

Lewis, in the very first page of his book[1]:

> "*If kangaroos had no tails, they would topple over*, seems to me to mean something like this: in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over."

We can observe that the word "resemble" may be seen as a reference to the concept of similarity between some possible state of affairs in relation to the actual state of affairs. The expression "as much as" here may be understood as a relative comparison of similarities among the possible states of affairs in relation to the actual state of affairs. But Lewis gave no formal definition of similarity in his book[1].

He defined two basic counterfactual conditional operators:

- $A \mathrel{\Box\!\!\rightarrow} B$: If it were the case that A, then it would be the case that B;
- $A \mathrel{\Diamond\!\!\rightarrow} B$: If it were the case that A, then it might be the case that B.

And provided also the definition of other counterfactual operators. But, since they are interdefinable, he took $\Box\!\!\rightarrow$ as the primitive for the construction of formulas.

In the middle of his book, he introduced the comparative possibility operators and showed that they can serve as the primitive notion for counterfactuals.

- $A \preccurlyeq B$: It is as possible that A as it is that B.

This operator gave us simpler proofs during this work.

He used possible-world semantics for intentional logic. For that reason the state of affairs are treated as worlds. To express similarity, he used proximity notions: a world is closer to the actual world in comparison to other worlds if it is more similar to the actual world than other considered worlds.

Lewis called the set of worlds to be considered for an evaluation as the strictness of the conditional. He pointed out that the strictness of the counterfactual conditional is based on the similarity of worlds. He showed that the counterfactual could not be treated by strict conditionals, necessity operators or possibility operators given by modal logics. To do so, he argued that strictness of the conditional can not be given before all evaluations. He constructed sequences of connected counterfactuals in a single English sentence for which the strictness cannot be given for the evaluation:

> ”If Otto had come, it would have been a lively party; but if both Otto and Anna had come it would have been a dreary party; but if Waldo had come as well, it would have been lively; but...”

to show that the strictness of the counterfactuals cannot be defined by the context, because the sentence provides a single context for the evaluation of all counterfactuals. If we try to fix a strictness that makes a counterfactual true, then the next counterfactual is made false.

Lewis proposed a variably strict conditional, in which different degrees of strictness is given for every world before the evaluation of any counterfactual. To express this concept, the accessibility relation is defined by a system of spheres, which is given for every world by a nesting function \$ that applies over a set of worlds $\mathcal{W}$. The nested function attributes a set of non-empty sets of worlds for each world and this set of sets is in total order for the inclusion relation.

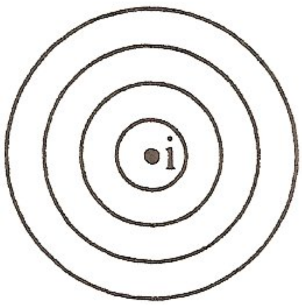

Figure I.1: A system of spheres around some world i

A systems of spheres, of any kind, is central in the most traditional analysis of counterfactuals. But the idea behind it is also available to many different logics. So, if we manage to handle them in a satisfactory manner, we will be able to use it in a broader class of logics. The system of neighbourhoods facilitates the development of the model, by leaving open the choice for a proper definition of similarity. And that concept can be used for a broader class of logics, not only the counterfactuals.

From Lewis definitions, the nesting function is a primitive notion:

> $\phi \mathrel{\Box\!\!\rightarrow} \psi$ is true at a world i (according to a system of spheres \$) if and only if either: no ϕ-world belongs to any sphere S in $\$_i$[1], or some sphere S in $\$_i$ does contains at least one ϕ-world, and $\phi \rightarrow \psi$ holds at every world in S.

> $\phi \preccurlyeq \psi$ is true at a world i (according to a system of spheres \$) if and only if, for every sphere S in $\$_i$, if S contains any ψ-world then S contains a ϕ-world.

Lewis[1] also provided conditions that may be applied to the nesting function \$. To every condition corresponds a different counterfactual logic:

- Normality (N): \$ is normal iff $\forall w \in \mathcal{W} : \$(w) \neq \emptyset$;
- Total reflexivity (T): \$ is totally reflexive iff $\forall w \in \mathcal{W} : w \in \bigcup \$(w)$;
- Weak centering (W): \$ is weakly centered iff $\forall w \in \mathcal{W} : \$(w) \neq \emptyset$ and $\forall N \in \$(w) : w \in N$;
- Centering (C): \$ is centered iff $\forall w \in \mathcal{W} : \{w\} \in \$(w)$;
- Limit Assumption (L): \$ satisfies the Limit Assumption iff, for any world w and any formula ϕ, if there is some ϕ-world[2] in $\bigcup \$(w)$, then there is some smallest sphere of $\$(w)$ that contains a ϕ-world;
- Stalnaker's Assumption (A): \$ satisfies Stalnaker's Assumption iff, for any world w and any formula ϕ, if there is some ϕ-world in $\bigcup \$(w)$, then there is some sphere of $\$(w)$ that contains exactly one ϕ-world;
- Local Uniformity (U-): \$ is locally uniform iff for any world w and any $v \in \bigcup \$(w)$, $\bigcup \$(w)$ and $\bigcup \$(v)$ are the same;
- Uniformity (U): \$ is uniform iff for any worlds w and v, $\bigcup \$(w)$ and $\bigcup \$(v)$ are the same;
- Local absoluteness (A-): \$ is locally absolute iff for any world w and any $v \in \bigcup \$(w)$, $\$(w)$ and $\$(v)$ are the same;
- Absoluteness (A): \$ is absolute iff for any worlds w and v, $\$(w)$ and $\$(v)$ are the same.

The $\boldsymbol{V}$-logic is the most basic counterfactual logic presented by Lewis[1], where "V" stands for variably strict conditional. If, for example, we accept the centering condition (C), then we have the $\boldsymbol{VC}$-logic. Lewis showed in his book

[1] $\$_i$ gives the neighbourhoods around the world i. They are the available strictness to evaluate counterfactuals at i.

[2] A ϕ-world is a world in which ϕ holds.

a chart of 26 non-equivalent $\boldsymbol{V}$-logics that arises from the combinations of the conditions.

We prefer to call the spheres as neighbourhoods, because they represent better the concept of proximity, which Lewis used to express similarity. The neighbourhoods provide a relative way to compare distance. The world that is contained in a neighbourhood is closer to the actual world than other world that is not contained in that same neighbourhood.

As far as we know, there is only one natural deduction system for the counterfactuals, which is given by Bonevac [13]. But his system is designed to deals with the $\boldsymbol{VW}$-logic, since it contains the rule of counterfactual exploitation ($\Box\!\!\rightarrow$E), which encapsulates the weak centering condition. His approach to define rules for the counterfactual operators provides a better intuition of the counterfactual logic. His systems is expressive enough to deal with modalities and strict conditionals. The labelling of world shifts using formulas make easier to capture the counterfactual mechanics.

We also found the work of Sano [14] which pointed out the advantages of using the hybrid formalism for the counterfactual logic. He presented some axioms and rules for the $\boldsymbol{V}_{\mathcal{HC}(@)}$-logic that extends the $\boldsymbol{V}$-logic of Lewis.

Another interesting reference is the article of Gent[12], which presents a new sequent- or tableaux-style proof system for V C. His work depends on the operator $[\![]\!]$ and the definition of signed formulas.

We recently found a sequent calculus, provided by Lellmann[25], that treats the $\boldsymbol{V}$-logic of Lewis and its extensions. Its language depends on modal operators, specially the counterfactual operators $\Box\!\!\rightarrow$ and $\Box\!\!\Rightarrow$ and the comparative possibility operator $\preccurlyeq$.

As far as we know, our deduction system is the only one dealing with Lewis systems in a general form, that is, without using modalities in the syntax and treating the most basic counterfactual $\boldsymbol{V}$-logic.