Counterfactuals Philosophical Rough Notes

Material Conditional and Strict Conditional

- Two candidate formalisations of counterfactual conditionals in MPL are "$\phi \rightarrow \psi$" (material conditional) and "$\Box(\phi \rightarrow \psi)$" (strict conditional).
- The material conditional is unsatisfactory as a formalisation of counterfactual conditionals because it obeys the **false antecedent** inference pattern (i.e. $\neg\phi \vDash_K \phi \rightarrow \psi$) and it obeys the **true consequent** inference pattern (i.e. $\psi \vDash_K \phi \rightarrow \psi$). Natural language counterfactual conditionals apparently do not obey these inference patterns. For example, in natural language, we do not infer from **"it is not snowing" that "if it is snowing, the Philosophical Logic exam would be cancelled"**, and we do not infer from **"the lecture is on Monday" that "if the lecture is rescheduled to Thursday, then the lecture is on Monday"**.
- Both the material conditional and the strict conditional are unsatisfactory because each obeys both **augmentation** (strengthening the antecedent) and **contraposition** (i.e. $\phi \rightarrow \psi \vDash_K \phi \wedge \psi \rightarrow \chi$, $\phi \rightarrow \psi \vDash_K \neg\psi \rightarrow \phi$, $\Box(\phi \rightarrow \psi) \vDash_K \Box(\phi \wedge \chi) \rightarrow \psi$, $\Box(\phi \rightarrow \psi) \vDash_K \Box(\neg\psi \rightarrow \neg\phi)$). Natural language counterfactual conditionals apparently do not obey these inference patterns. For example, in natural language, we do not infer from **"if the switch was flipped, then the light would be on" that "if the switch was flipped, and if the bulb had been removed, then the light would be on", and from "if it rained, it would not rain heavily" that "if it rained heavily, it would not rain".**

Stalnaker's Conditional

- **No importation** $\phi \rightarrow (\psi \rightarrow \chi) \Rightarrow (\phi \wedge \psi) \rightarrow \psi$, unlike with material conditional and strict conditional. This is a virtue because of such counterexamples to importation as **"if Bill had married Laura, then if he had married Hillary, he would have been happy", where if Bill had married Laura and Hillary, he would have become a public spectacle and been unhappy.**
- **No transitivity** $\phi \rightarrow \psi + \psi \rightarrow \chi \Rightarrow \phi \rightarrow \chi$, unlike with material conditional and strict conditional. This is a virtue because of such counterexamples to transitivity as **"if I had not been born, Mike would have been my parent's oldest child; if my parents never met, I would not have been born; so if my parents had never met, Mike would have been my parent's oldest child"**.
- **No transposition** $\phi \rightarrow (\psi \rightarrow \chi) \Rightarrow \psi \rightarrow (\phi \rightarrow \chi)$, unlike with material conditional (but not strict conditional). This is a virtue because of such counterexamples to transposition as **"if Bill had married Laura, then if Bill had married Hillary, then his wife's name would be Hillary; so if Bill had married Hillary, then if Bill had married Laura, his wife's name would be Hillary".**

Disjunctive Antecedents

- In neither Lewis's nor Stalnaker's system is $P\Box\!\!\rightarrow R$ is semantic consequence of $(P \vee Q)\Box\!\!\rightarrow R$. But **"if we had surrendered we would have been shot" seems to be a semantic consequence of "if we had surrendered or tried to run away we would have been shot".**
- This seems to be because "if we had surrendered or tried to run away we would have been shot" is better formalised as $(P\Box\!\!\rightarrow R) \wedge (Q\Box\!\!\rightarrow R)$. The naive formalisation of the English sentence is inappropriate, as it also is for such English sentences as "there ain't no cake". What "if we had surrendered or tried to run away we would have been shot" really means is not that "if we had surrendered we would have been shot and if we had tried to run away we would have been shot".

Lewis Against Stalnaker

- Lewis disputes Stalnaker's assumption of antisymmetry (more precisely, Stalnaker's construction of an antisymmetric nearness relation). The simple thought is that, in general, two things can be equally similar to one other thing, so **there seems to be no reason to think that two possible worlds could not be equally near to one other possible world.** Lewis objects to the assumption of antisymmetry on metaphysical grounds. If the truth of a natural language counterfactual does in fact depend on some nearness relation, then simply because there could be equally near possible worlds, the assumption of antisymmetry is inappropriate for natural language counterfactuals.
    - **Stalnaker concedes that the antisymmetry assumption is an implausible assumption about the kind of nearness relation we use to interpret English counterfactual conditionals.** But Stalnaker defends this assumption as an **"idealising assumption"**, akin to the assumptions (made in other languages of logic or semantic

theories) that domains, sets of possible worlds, and predicates are assumed to have sharp boundaries, even though in natural language there are borderline cases. In other words, in English, the domains of discourse are vague, and there are borderline cases, yet it does not trouble us that in, for example, first-order predicate logic, the domain of discourse is treated as well-defined. **That Stalnaker's logic of counterfactuals is determinate while there are cases of indeterminacy in reality, is an instance of the general problem of the vagueness of natural language ("pervasive semantic underdetermination in natural language"), rather than a fault of Stalnaker's antisymmetry assumption.**

- **Conditional excluded middle is true under Stalnaker's semantics.** Stalnaker considers this to be a virtue. An equivalent formulation (that simply deconstructs $\vee$ into $\neg$ and $\rightarrow$) of conditional excluded middle is $\neg(\phi \mathrel{\square\!\!\rightarrow} \psi) \rightarrow (\phi \mathrel{\square\!\!\rightarrow} \neg\psi)$. Where ϕ is possibly true, the converse, $(\phi \mathrel{\square\!\!\rightarrow} \neg\psi) \rightarrow \neg(\phi \mathrel{\square\!\!\rightarrow} \psi)$ is uncontroversially true. **So if conditional excluded middle is valid, then where ϕ is possibly true, $\neg(\phi \mathrel{\square\!\!\rightarrow} \psi)$ and $\phi \mathrel{\square\!\!\rightarrow} \neg\psi$ are logically equivalent. We indeed treat the English analogues as logically equivalent.** For example, we do not distinguish between "it's not true that if she had played she would have won" and "if she had played she would not have won".
  - **Lewis considers this to be a fault**. In the **Bizet and Verdi** case, it seems neither true that if the two were compatriots, Bizet would be Italian, nor true that if the two were compatriots, Bizet would not be Italian. (Though it is certainly true that if the two were compatriots, Bizet would either be Italian or not be Italian.)
  - Stalnaker argues that, because of "pervasive semantic underdetermination in natural language", **the solution is not to amend Stalnaker's semantics, but to introduce some theory of vagueness and employ some method for handling indeterminacy (it is indeterminate whether an Italian-Bizet-world or a non-Italian-Bizet-world is closer)** like supervaluationism. Stalnaker's semantic theory, **appropriately augmented by supervaluation, would consider both the statements neither true nor false rather than flatly false. And this seems more natural, most speakers would be as hesitant to deny as to affirm either of the conditionals.**
- **Distribution is valid under Stalnaker's semantics.** Stalnaker considers this to be a virtue. If someone says "if I had been a baseball player, I would have been either a third-baseman or a shortstop", it seems natural to reply "well which would you have been?" This reply seems to presuppose that either (1) if this person had been a baseball player, he would have been a third-baseman, or (2) if this person had been a baseball player he would have been a shortstop.
  - **Lewis considers this to be a fault. According to Lewis, the potential baseball player neither would have been a third-baseman nor would have been a shortstop, but might have been either.** The question does not presuppose either (1) or (2), because it makes sense to ask the question even if one anticipates the response "I would have been one of the two for sure, but it could really be either one".
- **Lewis objects that Stalnaker's semantics cannot accommodate a might counterfactual.** Under Lewis's definition of the might counterfactual, "$\phi \mathrel{\diamond\!\!\rightarrow} \psi$" abbreviates "$\neg(\phi \mathrel{\square\!\!\rightarrow} \neg\psi)$", the might counterfactual and the would counterfactual are equivalent under Stalnaker's semantics. Lewis considers and rejects alternative formalisations of the might counterfactual in Stalnaker's semantics. Notably, the formalisation $\diamond(\phi \mathrel{\square\!\!\rightarrow} \psi)$.
  - Stalnaker interprets a might counterfactual as "it might be the case that if ... then ..." where **might expresses epistemic possibility**. This has the advantage over Lewis's interpretation that it **does not treat "if ... might ..." as an idiom, but rather as analysable in terms of its direct components**. That "might" is taken to apply not only to the consequent but to the conditional is not ad hoc because we do the same in "if he is a bachelor, he must be unmarried". In English, "might" generally expresses epistemic possibility (though not always). Then, Stalnaker has a response to Lewis's Penny case.
- Lewis objects to the limit assumption on the ground of the **one inch line case, where it seems there is no actual-world-closest more-than-one-inch-world.**
  - For Stalnaker, the selection function is primitive and the nearness function is induced by the selection function, so the assumption that some "nearest" world exists is not arbitrary but implied by the way the nearness relation is constructed. Admittedly, an independently understood intuitive notion of similarity among possible worlds is necessary to explain the basis of a selection. Stalnaker's intuitive explanation is that the selection function selects a world that makes a minimal change in the relevant respects to the actual world, such that the antecedent is true. So in Lewis's one inch line case, the selection function would simply select some world in which the line was longer than and reasonably close to one inch. Presumably, it remains indeterminate which such world would be selected, but this problem of indeterminacy is to be solved by a theory of vagueness and an appropriate "method" such as supervaluation.