The Structure of Possible Worlds

05.31.12

Will Starr – Cornell University (Philosophy)

UCLA Semantics Workshop

1 Background

- Counterfactual conditionals, like (1), can be false

  (1) If Kennedy hadn't been assassinated, he would have grown wings

  - But the corresponding material conditional $\neg\mathsf{A} \supset \mathsf{W}$ can't be
  - Indeed, since some counterfactuals are also true, no truth-functional analysis is possible, e.g. *if Kennedy hadn't been assassinated, Jackie O would have left Dallas happy*
- Goodman (1947) introduced the co-tenablity analysis of counterfactuals but tempered it with some despair about any analysis

  Co-tenability Analysis A counterfactual $\mathsf{A} > \mathsf{B}$ is true just in case B follows from all facts which are co-tenable with A.

  - 'Co-tenability' can't be logical consistency, though counterfactually assuming A does require giving up $\neg\mathsf{A}$
    - But it may require giving up things that followed by law from $\neg\mathsf{A}$.
      - ▷ (The switch is down and the light is off) If the switch had been flipped, the light would be off
    - Also, assuming A may require adding things that follow by law from it.
      - ▷ (The switch is down and the light is off) If the switch had been flipped, the room would not be dark
  - Goodman's 1st despair: what are laws?
    - Can we characterize them in any way other than as true counterfactuals?
    - Can we characterize them without relying on 'unscientific' concepts like causation?
    - 'Following from laws' is unlike deduction: from *the match was struck* it follows by law that *it lit*, but from *the match was struck and it was soaked in water* it does not follow by law that *it lit* ('non monotonic')
  - Goodman's 2nd despair: the facts are hard to spell out
    - Unless the facts themselves are non-counterfactual, the analysis is circular
    - Knowing a counterfactual requires knowing all non-counterfactual facts!
- Then came analyses in terms of *possible worlds*
  - Intuitively: a complete and consistent way the world could be
  - Formally: points at which all atomic formulas have exactly one truth value, and which are related by 'accessibility' R
    - Modal logic: $[\![\Box\phi]\!] = \{w \mid R(w) \subseteq [\![\phi]\!]\}$, $[\![\mathsf{A}]\!] = \{w \mid v(w, \mathsf{A}) = 1\}$
- Stalnaker and Lewis used *similarity* rather than *accessibility*

Lewis-Stalnaker Semantics (Stalnaker 1968; Stalnaker & Thomason 1970; Lewis 1973)

- $\phi > \psi$ is true at w just in case all of the ϕ-worlds most **similar** to w are ψ-worlds
  - Most similar according to the selection function f
  - f takes a proposition p and a world w and returns the p-worlds most similar to w
- $[\![\phi > \psi]\!]^f = \{w \mid f(w, [\![\phi]\!]^f) \subseteq [\![\psi]\!]^f\}$
  - Similarity, unlike accessibility, is non-monotonic: $f(w, p \cap q) \nsubseteq f(w, p)$
    - Yet: $R(w) \cap (p \cap q) \subseteq R(w) \cap p$

(Making the 'Limit Assumption': there are most similar worlds)

- How does this avoid Goodman's puzzles?

  Lewis-Stalnaker on Goodman 'Abstract Analysis': We can say enough about similarity to capture the *logic* of counterfactuals without answering Goodman's questions.

  - By placing a few constraints on which worlds count as most similar, Lewis and Stalnaker are able to get a plausible (though debated) logic for counterfactuals
  - Some of these constraints are motivated by the intuitive concept of similarity itself
  - Some are imposed just to get the right logic
  - Logic of $>$ is determined by constraints on f (where $p, q \subseteq W$ and $w \in W$):

(a)	$f(w,p) \subseteq p$	**success**
(b)	$f(w,p) = \{w\}$, if $w \in p$	**strong centering**
(c)	$f(w,p) \subseteq q$ & $f(w,q) \subseteq p \implies f(w,p) = f(w,q)$	**uniformity**
(d)	$f(w,p)$ contains *at most* one world	**uniqueness**

  Stalnaker Constraints (a)-(d)

  Limited Lewis Constraints (a)-(c)

  - 'Limited Lewis': Lewis if he had accepted the Limit Assumption

Email: will.starr@cornell.edu.
URL: http://williamstarr.net.

2 Challenging the Abstract Analysis

- Communication problems:
  - A proposition is conveyed w/$\mathsf{A} > \mathsf{B}$ only when a particular f is filled in
    - There are thousands of f's meeting the formal constraints
  - Stalnaker and Lewis regard this filling in of f as a standard case of context sensitivity
    - Context sensitivity is the process of using information mutually available in the utterance situation to interpret an utterance
    - More precisely: interpretations of utterances that would fail to communicate anything without using information mutually available in the context
  - We must all therefore have the means for getting f's from available information
  - Lewis and Stalnaker: ordinary concept of similarity serves this role
  - **Issue 1**: filling in f seems to require solving Goodman's puzzles
    - Maybe there is so much vagueness in f, this is an instance of the more general puzzle of how communication with vague language works
  - **Issue 2**: the facts that determine similarity make uttering the subjunctive redundant
    - Rather than taking f as fixed and communicating something on the basis of it, subjunctives seem to inform us about f
    - This doesn't mesh with the standard model of context sensitivity
  - **Issue 3**: various examples demonstrate that it is not our intuitive concept of similarity that is put to use in determining the truth value of subjunctive conditionals
- Abstract as it is, the similarity analysis sometimes still predicts the wrong truth-conditions
- The future-similarity problem (Fine 1975: 452):

  (2) If Nixon had pressed the button there would have a been a nuclear holocaust
  $\mathsf{B} > \mathsf{H}$

  - Plausibly, (2) is true, or can at least supposed to be.
  - Suppose further that there never will be a nuclear holocaust.
  - Then, for every $\mathsf{B} \wedge \mathsf{H}$-world, there will be a closer $\mathsf{B} \wedge \neg\mathsf{H}$-world
    - In this world a small change prevents the holocaust, such as a malfunction in the electrical detonation system
  - The idea: surely a world where Nixon presses the button and a malfunction prevents nuclear holocaust is more like our own than one where there is a nuclear holocaust!
  - So it would seem that the Lewis-Stalnaker theory predicts (2) to be false!

- The fact-fixing problem (Tichý 1976: 271):

  (3) a. Invariably, if it is raining, Jones wears his hat
  b. If it is not raining, Jones wears his hat at random
  c. Today, it is raining and so Jones is wearing his hat
  d. But, even if it had't been raining, Jones would have been wearing his hat

  - Given (3a-c), (3d) seems clearly incoherent/false/bad
  - Why is this a counterexample to the Stalnaker-Lewis theory?
    - In the actual world $w_@$, Jones is wearing his hat.
    - So in the non-raining-worlds most similar to $w_@$, Jones is wearing his hat
    - But then Stalnaker-Lewis predict that (3d) is true!
- Lewis (1979) articulated a system of weights for similarity which was, among other things, supposed to address these counterexamples:

  (1) First importance: avoid big, widespread, diverse violations of law. ('big miracles')

  (2) Second importance: maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
    - Maximize match in matters of fact before 'miracles' occur

  (3) Third importance: avoid even small, localized, simple violations of law. ('little miracles')

  (4) Little or no importance: secure approximate similarity of particular fact, even in matters that concern us greatly.
- This system works for the two counterexamples by not counting particular matters of fact towards similarity
- As Lewis (1979: 466-7) acknowledged, this gives up the idea that intuitive similarity is involved in evaluating subjunctive conditionals
  - But this is a problem!
- Further the system of weights doesn't cover a simple variant of Tichy's case:

  (4) a. Before Jones opens the curtain to see what the weather is like, he flips a coin
  b. If it's not raining and the coin comes up heads, he wears his hat
  c. If it's not raining and the coin comes up tails, he doesn't wear his hat
  d. Invariably, if it is raining he wears his hat
  e. Today, the coin came up heads and it is raining, so Jones is wearing his hat
  f. But, even if it hadn't been raining, Jones would have been wearing his hat

  (Veltman 2005: 164)
  - Here, when you counterfactually suppose that it isn't raining, you *do* keep fixed the subsequent outcome of the coin toss and the hat wearing
  - So particular matters of fact are sometimes kept fixed, even after the 'miracle' occurs

- A simpler case that illustrates the same phenomenon:

  (5) [Suppose there is a circuit such that the light is on exactly when both switches are up. A kid is playing with the switches before we enter the room. He flicks switch one down and then switch two up, so the lamp is out.] When we walk in, I say: If switch one had been up, the lamp would have been on.
  (Lifschitz)

  - We keep fixed the fact that switch two is up, even though it is after the 'miracle' necessary to flip up switch one!
  - There are many more like this...
- Lewis (1979: 472) notes that in cases like these, things come out 'differently' and 'would like to know why'
- Plausible diagnosis (Veltman 2005: 164):

  > Similarity of particular fact is important, but only for facts that do not **depend** on other facts. Facts stand and fall together. In making a counterfactual assumption, we are prepared to give up everything that depends on something that we must give up to maintain consistency. But, we want to keep in as many independent facts as we can.

  - In Tichy's case, when we counterfactually suppose that it isn't raining we don't keep fixed the fact that he is wearing his hat because his wearing his hat depended on the fact that it is raining
  - In the variant, the outcome of the coin flip was independent of raining, so we keep it fixed when we suppose that it wasn't raining
    - But then it follows, because of (4b), that Jones would be wearing his hat
- Kratzer (1989) and Veltman (2005) propose an analysis of counterfactuals that is sensitive to **dependence**
  - Couched in a *situation semantics*
  - Situations: consistent sets of facts, worlds are maximal situations
  - Basic idea: q depends on p just in case, for any situations s, s': q is true in s' if p is true in s and $s \subset s'$
    - Distributional analysis of dependence
  - Sketch of semantics: no.
- As Schulz (2007: 101) notes, neither theory makes the right prediction for (5)
  - Evidencing that the distributional analysis of dependence is not quite right
- Schulz (2011, 2007: §5.6) and Hiddleston (2005) develop analyses of counterfactuals inspired by the analysis given in Pearl (1995, 2000: Ch.7)
  - Based on the causal models and structural equations pioneered by Pearl (1993) and Spirtes *et al.* (1993)

  - It offers great promise in correctly capturing our cases while also capturing the logic of counterfactuals
- I'll pursue this strategy too, but note the bigger picture
  - We are back to directly answering Goodman's puzzles, rather than attempting an 'abstract analysis'
- Is this a good thing?
  - If we can't answer Goodman's puzzles, then we can't answer some of the most basic questions about what the world is like and how we can have knowledge of it
  - When a robot with some knowledge about its environment performs an action that changes that environment, how does the robot decide to update its knowledge to reflect this change? (McCarthy & Hayes 1969)
  - Striking the match tends to lead to fire, but not in some circumstances, like when it's wet, or when the striking surface is too smooth, or when there's no oxygen, or when the robot is confusing a twig for a match and so on.
  - This is the *frame problem* (Shanahan 2009)
  - Pearl's example:
    - *Input*:
      (1) Suitcases open iff both locks are open
      (2) The right lock is open
      (3) The suitcase is closed
    - *Query*:
      (1) What would happen if we open the left lock?
      (2) The right lock might get closed
- Perhaps now it is clear why a computer scientist would care about these issues!
- The structural equations approach has lots of nice peripherals too
  - Work on how it can be used to formulate a theory of explanation: Woodward (2003), Halpern & Pearl (2005a,b)
  - Work on how causal models can be inferred from observation Pearl (2000)

3 Causal Models and Structural Equations

- In standard propositional logics, atomic sentences are assigned independent truth-values
  - A valuation v, is a simple function from atomics to truth-values
    - E.g. $v(\mathsf{A}, w) = 1, v(\mathsf{B}, w) = 0, \ldots$
- Pearl's causal models give up this assumption
  - The truth-value of an atomic D can **depend** on the truth-value others A and B
- These dependencies are *functional*
  - If D depends only on A and B, then Ds' truth-value is uniquely determined by D and A
- One of Pearl's most important contributions is making these models probabilistic
  - But we're not going to going to go into that
- Following Simon (1953), Pearl thinks of dependencies as invariant 'causal mechanisms'
  - Like Simon, he represents them out as a set of 'structural equations'
- D's truth depends on both A and C being true, $\mathsf{D} := \mathsf{A} \wedge \mathsf{C}$
  - Or D's truth depends on one of them being true, $\mathsf{D} := \mathsf{A} \vee \mathsf{C}$
- You can picture the models underlying these equations as directed graphs

3.1 Pearl's (2002) Prisoner Example

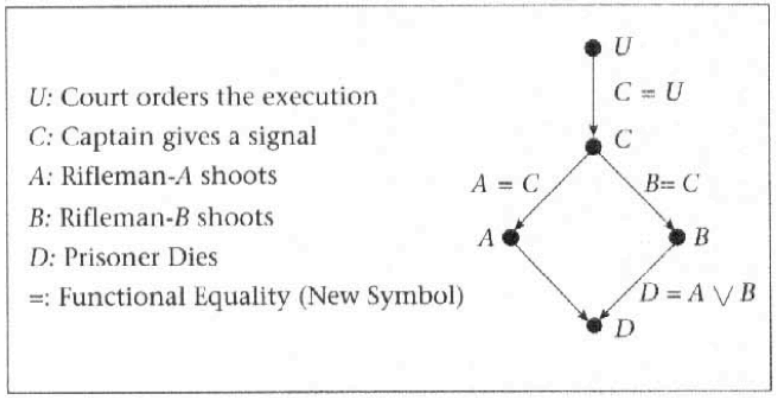

Figure 4. Causal Models at Work (the Impatient Firing Squad).

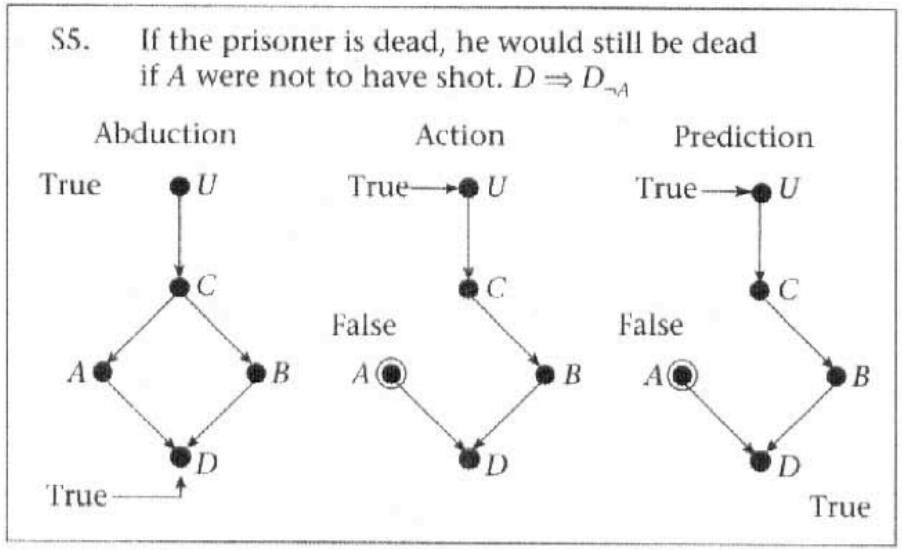

Figure 7. Three Steps to Computing Counterfactuals.

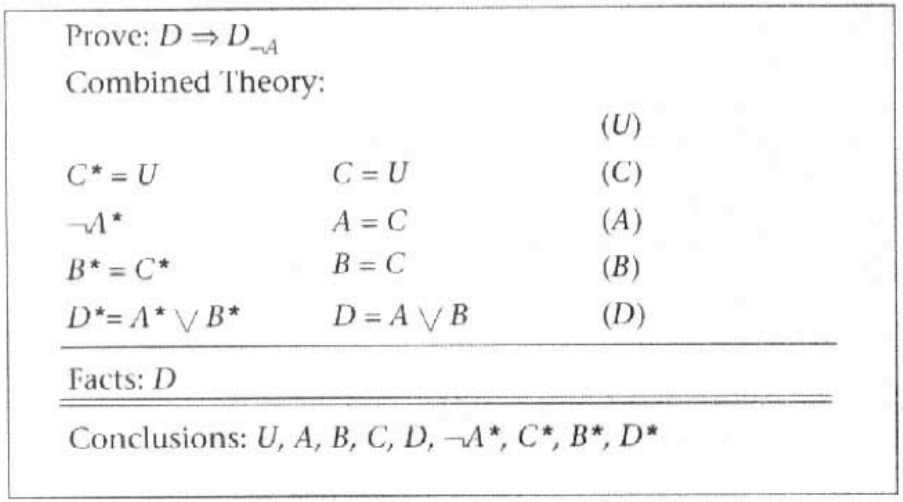

Figure 8. Symbolic Evaluation of Counterfactuals.

- The account of the light example works pretty much the same!
  - Instead of keeping fixed that B shot, we keep fixed that switch two is up

4 Structural Equations and the Structure of Possible Worlds

- Pearl's approach is clearly promising, but existing implementations have limitations
  - Exogenous atomic sentences (ones with no arrows going into them) cannot be manipulated by 'intervention': if the order hadn't been given, prisoner would be alive!
  - Logically complex antecedents are problematic; consequents that are counterfactuals
  - It would be nice to be able to write equations in the antecedents to capture counterlegals
    - If switch one alone controlled the light, then the light would be on
  - Not all counterfactuals are about causal connections
    - If this cup were red, it wouldn't be blue
  - It does not easily integrate with the existing tools of formal semantics
- My contribution: a formalism without these limitations
- Starting point: classical possible worlds are valuations (situations are partial valuations)
  - Worlds fix the truth values of each atomic sentence
  - Picture each atomic as a dot, which is black if false, white if true.

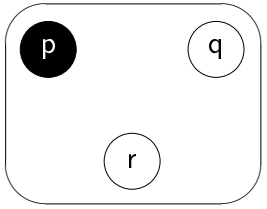

Fig. 1. Classical possible world w

$$w(\mathsf{p}) = 0$$
$$w(\mathsf{q}) = 1$$
$$w(\mathsf{r}) = 1$$

Fig. 2. System of equations for w

- Now depart from the classical picture (Starr 2012):
  - The dependencies between facts endow worlds with a structure

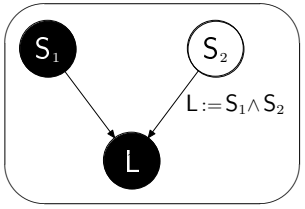

Fig. 3. A structured possible world w

$$w(\mathsf{S_1}) = 0 \quad (6)$$
$$w(\mathsf{S_2}) = 1 \quad (7)$$
$$w(\mathsf{L}) = w(\mathsf{S_1}) \cdot w(\mathsf{S_2}) \quad (8)$$
$$= 0$$

Fig. 4. Equations for w

- $\neg, \wedge$ and $\vee$ all have arithmetic counterparts operating on 1 and 0

$\neg$	$\wedge$	$\vee$
$1 - x$	$x \cdot y$	$(x + y) - (x \cdot y)$

- To evaluate the counterfactual $\mathsf{S_1} > \mathsf{L}$, create world $w_{\mathsf{S_1}}$
  - Step 1: intervention
    - Eliminate old assignment for $\mathsf{S_1}$, line (9)
    - Make $\mathsf{S_1}$ 1, line (10)
  - Step 2: projection
    - Apply equation (12) to solve for L
    - New result: $w_{\mathsf{S_1}}(\mathsf{L}) = 1$!

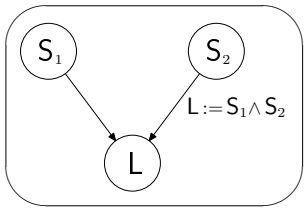

Fig. 5. The New World $w_{\mathsf{S_1}}$

$$w(\mathsf{S_1}) = 0 \quad (9)$$
$$w_{\mathsf{S_1}}(\mathsf{S_1}) = 1 \quad (10)$$
$$w_{\mathsf{S_1}}(\mathsf{S_2}) = w(\mathsf{S_2}) = 1 \quad (11)$$
$$w_{\mathsf{S_1}}(\mathsf{L}) = w(\mathsf{L}) = w_{\mathsf{S_1}}(\mathsf{S_1}) \cdot w_{\mathsf{S_1}}(\mathsf{S_2}) \quad (12)$$
$$= 0$$
$$= 1 \quad (13)$$

Fig. 6. Equations for $w_{\mathsf{S_1}}$

- To see how this works better, consider a slightly modified scenario:
  - Switch 1 turns on a servo that controls switch 2: $\mathsf{S_2} := \mathsf{S_1}$
  - Switch 2 turns on the light: $\mathsf{L} := \mathsf{S_2}$
  - Currently, switch 1 is up, so 2 is up and the light is on

  (14) If switch 2 were up, the light would be off
- This comes out true, because after setting $\mathsf{S_2}$ to 1, the equation connecting it will make L come out as 1 too
  - Lesson: you only keep fixed facts which are not determined by facts you are counterfactually giving up
- These interventions are quite like Lewis' miracles
  - We go to a world exactly like w except w's mechanisms have been broken to allow a particular fact to hold, which ends up generating a world very dissimilar to w
- This captures Fine's Nixon case nicely.

4.1 What are Worlds?

- The idea spelled out for w:
  - Each independent atomic is mapped to a truth-value: $\{\langle \mathsf{S}_2, 1\rangle, \langle \mathsf{S}_1, 0\rangle\}$
  - Pair L with a dependency function d that determines the truth of L from a pairing of $\mathsf{S}_1, \mathsf{S}_2$ with truth-values:

S_1	S_2	L
1	1	1
1	0	0
1	0	0
1	1	0

    - d maps L and $\{\langle \mathsf{S}_1, 1\rangle, \langle \mathsf{S}_2, 1\rangle\}$ to 1
    - d maps L and $\{\langle \mathsf{S}_1, 1\rangle, \langle \mathsf{S}_2, 0\rangle\}$ to 0
    - d maps L and $\{\langle \mathsf{S}_1, 0\rangle, \langle \mathsf{S}_2, 1\rangle\}$ to 0
    - d maps L and $\{\langle \mathsf{S}_1, 0\rangle, \langle \mathsf{S}_2, 0\rangle\}$ to 0
- So, a world is a function from some independent atomics to truth-values, together with a dependency function for each dependent atomic
- $w = \{\langle \mathsf{S}_2, 1\rangle, \langle \mathsf{S}_1, 0\rangle, d\}$
  - $d = \{ \quad \langle \langle \mathsf{L}, \{\langle \mathsf{S}_1, 1\rangle, \langle \mathsf{S}_2, 1\rangle\}\rangle, 1\rangle,$
    $\langle \langle \mathsf{L}, \{\langle \mathsf{S}_1, 1\rangle, \langle \mathsf{S}_2, 0\rangle\}\rangle, 0\rangle,$
    $\langle \langle \mathsf{L}, \{\langle \mathsf{S}_1, 0\rangle, \langle \mathsf{S}_2, 1\rangle\}\rangle, 0\rangle,$
    $\langle \langle \mathsf{L}, \{\langle \mathsf{S}_1, 0\rangle, \langle \mathsf{S}_2, 0\rangle\}\rangle, 0\rangle \quad \}$
- Dependencies are functions from pairs of atomics and situations to truth-values

4.2 Extending the Analysis

- We'd like to define the notation w_ϕ for any non-counterfactual ϕ
- Let's first try just with compounds of atomics:
  - w_{A} is the world exactly like w except that it assigns A to 1
  - $w_{\neg \mathsf{A}}$ is the world exactly like w except that it assigns A to 0
  - $w_{\mathsf{A} \wedge \mathsf{B}}$ is the world exactly like w except that it assigns A and B to 1
  - $w_{\mathsf{A} \vee \mathsf{B}}$ is the world exactly like w except that it assigns ?????????
- An idea: allow w_ϕ to be a *set* of worlds
  - w_{A} are the worlds exactly like w except that they assign A to 1
  - $w_{\neg \mathsf{A}}$ are the worlds exactly like w except that they assign A to 0
  - $w_{\mathsf{A} \wedge \mathsf{B}} = (w_{\mathsf{A}})_{\mathsf{B}} = \{w' \mid \exists w'' \in w_{\mathsf{A}} \ \& \ w' \in w''_{\mathsf{B}}\}$
    - Program sequencing '$\alpha\,;\beta$' from Dynamic Logic (Harel *et al.* 2000)
  - $w_{\mathsf{A} \vee \mathsf{B}} = w_{\mathsf{A}} \cup w_{\mathsf{B}}$
    - Program choice '$\alpha \cup \beta$' from Dynamic Logic Harel *et al.* (2000)

- Attempt to generalize:
  - $w_{\phi \wedge \psi} = (w_\phi)_\psi = \{w' \mid \exists w'' \in w_\phi \ \& \ w' \in w''_\psi\}$
  - $w_{\phi \vee \psi} = w_\phi \cup w_\psi$
  - $w_{\neg \phi}$ are the worlds exactly like w except that they assign....????
- Drawing inspiration from dynamic logic negation, converse '$\alpha^{\leftarrow}$':
  - Tempting: $w_{\neg \phi} = W - w_\phi$
    - But wrong: this will contain some ϕ-worlds and $\neg\phi$-worlds not related to w by a minimal change
  - Different idea: look for the $\neg\phi$-worlds which, with a minimal change, would become w
  - Consider each $\neg\phi$-world w'. If $w \in w'_\phi$, then $w' \in w_{\neg\phi}$.
    - If w is a $\neg\phi$-world, then $\{w\} = w_{\neg\phi}$

Dependency Semantics for Counterfactuals

- $[\![\phi > \psi]\!] = \{w \mid w_\phi \subseteq [\![\psi]\!]\}$
- $\phi > \psi$ is true iff either ψ is independent of ϕ and true, or else ϕ is sufficient for bringing about ψ when holding fixed all those facts that do not depend upon ϕ.[1]
  - Entailment, truth, etc. is all defined classically

4.3 Remaining Issues

- Comparison with Briggs (to appear)?
  - Briggs notes that on her formalization, and Pearl's, modus ponens fails for:

    (15) If executioner A had fired, then (even) if the captain had not signalled, the prisoner would have died.
    - Consider the whole conditional in a world where A did fire, which has the structure in Fig. 4
    - Pearl/Briggs semantics: make A true by intervention (erase all incoming lines to A), make C false by intervention. The line between A and D still means that D will be true.
    - Now consider the consequent conditional in the original world. The line between C and A will still be there, so D will come out false. But then the consequent conditional is false!
  - On mine, it does not.
    - We are evaluating (15) in a world w where A is true, so $w_{\mathsf{A}} = w$

[1] This intuitive paraphrase is from Cumming (2009: 1).

- ▶ So (15) just comes out false.

- What about backtrackers?
  - If the light had been on, the two switches would have to have been up
  - The very beginnings of a story: consider worlds where the light is on
    - ▶ No intervention will be necessary to turn it on
    - ▶ Render the consequent as: the set of all worlds w, s.t. $\mathsf{S}_1 \wedge \mathsf{S}_2$ is a necessity among all of the no-intervention L-worlds (vacuous quantification over w)
    - ▶ Perhaps this is a plausible semantics of the extra *have to*
    - ▶ Recall: $[\![\phi > \psi]\!] = \{w \mid w_\phi \subseteq [\![\psi]\!]\}$
    - ▶ If the necessity holds then $[\![\mathsf{Have.to}\,(\mathsf{S}_1 \wedge \mathsf{S}_2)]\!] = W$, so the conditional is true
    - ▶ If the necessity fails, then $[\![\mathsf{Have.to}\,(\mathsf{S}_1 \wedge \mathsf{S}_2)]\!] = \varnothing$, so the conditional is false
- What about counter-legals?
  - Consider: $(\mathsf{L} := \mathsf{S}_1 \vee \mathsf{S}_2) > \mathsf{L}$
    - ▶ $(\mathsf{L} := \mathsf{S}_1 \vee \mathsf{S}_2)$ denotes a dependency d
      - ▷ $d = \{ \quad \langle \langle \mathsf{L}, \{\langle \mathsf{S}_1, 1\rangle, \langle \mathsf{S}_2, 1\rangle\}\rangle, 1\rangle,$
        $\langle \langle \mathsf{L}, \{\langle \mathsf{S}_1, 1\rangle, \langle \mathsf{S}_2, 0\rangle\}\rangle, 1\rangle,$
        $\langle \langle \mathsf{L}, \{\langle \mathsf{S}_1, 0\rangle, \langle \mathsf{S}_2, 1\rangle\}\rangle, 1\rangle,$
        $\langle \langle \mathsf{L}, \{\langle \mathsf{S}_1, 0\rangle, \langle \mathsf{S}_2, 0\rangle\}\rangle, 0\rangle \quad \}$
  - We can then define $w_{\mathsf{A}:=\phi}$ as: w with $\langle \mathsf{A}, x\rangle$ removed and the dependency denoted by $\mathsf{A} := \phi$ added in its place
- Consider a case where one switch controls a light; the switch is up and the light on:
  - But if the light had been off, then if you had flipped the switch up, the light would have come on
- What properties must dependencies have to ensure formulas always end up with a truth value and never end up truth-valueless?
- Quantifiers and predicate logic?

References

Briggs, R (to appear). 'Interventionist Counterfactuals.' *Philosophical Studies*. URL http://www.rachaelbriggs.net/Rachael_Briggs/CV_(with_online_papers)_files/CM_Counterfac_6.pdf.

Cumming, S (2009). 'On What Counterfactuals Depend.' Ms. UCLA.

Fine, K (1975). 'Review of Lewis' *Counterfactuals*.' *Mind*, **84**: 451–8.

Goodman, N (1947). 'The Problem of Counterfactual Conditionals.' *The Journal of Philosophy*, **44**: 113–118.

Halpern, J & Pearl, J (2005a). 'Causes and Explanations: A Structural-Model Approach. Part I: Causes.' *British Journal for Philosophy of Science*, **56**.

Halpern, J & Pearl, J (2005b). 'Causes and Explanations: A Structural-Model Approach. Part II: Explanations.' *British Journal for Philosophy of Science*, **56**: 889–911.

Harel, D, Kozen, D & Tiuryn, J (2000). *Dynamic Logic*. Cambridge, MA: MIT Press.

Hiddleston, E (2005). 'A Causal Theory of Conditionals.' *Noûs*, **39(4)**: 632–657.

Kratzer, A (1989). 'An Investigation of the Lumps of Thought.' *Linguistics and Philosophy*, **12(5)**: 607–653.

Lewis, DK (1973). *Counterfactuals*. Cambridge, Massachusetts: Harvard University Press.

Lewis, DK (1979). 'Counterfactual Dependence and Time's Arrow.' *Noûs*, **13**: 455–476.

McCarthy, J & Hayes, PJ (1969). 'Some Philosophical Problems from the Standpoint of Artificial Intelligence.' In B Meltzer & D Michie (eds.), *Machine Intelligence 4*, 463–502. Edinburgh: Edinburgh University Press.

Pearl, J (1993). 'Graphical Models, Causality and Intervention.' *Statistical Science*, **8(3)**: 266–273.

Pearl, J (1995). 'Causation, Action, and Counterfactuals.' In A Gammerman (ed.), *Computational Learning and Probabilistic Learning*. New York: John Wiley and Sons.

Pearl, J (2000). *Causality: Models, Reasoning, and Inference*. Cambridge, England: Cambridge University Press.

Pearl, J (2002). 'Reasoning with Cause and Effect.' *AI Magazine*, **23(1)**: 95–112. URL http://ftp.cs.ucla.edu/pub/stat_ser/r265-ai-mag.pdf.

Schulz, K (2007). *Minimal Models in Semantics and Pragmatics: free choice, exhaustivity, and conditionals*. Ph.D. thesis, University of Amsterdam: Institute for Logic, Language and Information, Amsterdam. URL http://www.illc.uva.nl/Publications/Dissertations/DS-2007-04.text.pdf.

Schulz, K (2011). 'If you'd wiggled A, then B would've changed.' *Synthese*, **179**: 239–251. 10.1007/s11229-010-9780-9, URL http://dx.doi.org/10.1007/s11229-010-9780-9.

Shanahan, M (2009). 'The Frame Problem.' In EN Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, winter 2009 edn. URL http://plato.stanford.edu/archives/win2009/entries/frame-problem/.

Simon, HA (1953). 'Causal Ordering and Identifiability.' In WC Hood & TC Koopmans (eds.), *Studies in Econometric Method*, 49–74. New York: Wiley.

Spirtes, P, Glymour, C & Scheines, R (1993). *Causation, Prediction, and Search*. Berlin: Springer-Verlag.

Stalnaker, RC (1968). 'A Theory of Conditionals.' In N Rescher (ed.), *Studies in Logical Theory*, 98–112. Oxford: Basil Blackwell Publishers.

Stalnaker, RC & Thomason, RH (1970). 'A Semantic Analysis of Conditional Logic.' *Theoria*, **36**: 23–42.

Starr, WB (2012). 'The Structure of Possible Worlds.' Ms. Cornell University.

Tichý, P (1976). 'A Counterexample to the Stalnaker-Lewis Analysis of Counterfactuals.' *Philosophical Studies*, **29**: 271–273.

Veltman, F (2005). 'Making Counterfactual Assumptions.' *Journal of Semantics*, **22**: 159–180. URL http://staff.science.uva.nl/~veltman/papers/FVeltman-mca.pdf.

Woodward, J (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.