NASSLLI 2016—Multi-Modal Logic

Reinhard Muskens

Tilburg Center for Logic, Ethics, and Philosophy of Science (TiLPS)

Part IV: Conditional Modalities

Conditional Modalities

- We will now have a look at logics for counterfactuals, conditional obligation, 'soft' belief and the like.
- These logics have in common that they are based on preference relations, possibly in conjunction with some (other) accessibility relations. They have a lot of overlap from a technical point of view.
- We will look at counterfactuals first; at other conditional modalities later.
- We will base ourselves on Robert Stalnaker's *Theory of Conditionals* (1968), David Lewis's *Counterfacuals* (Blackwell 1973) and later work by Burgess and Boutilier. The tableau rules are new (but follow straightforwardly from the ideas expressed by some of these authors).

Counterfactual Conditionals—Some Examples

- If kangaroos had no tails, they would topple over. (Lewis, 1973)
- If we had all been living in California, things would have been different.
- If Granny were still alive, she would have loved the whisky-tasting.

We formalise this with the help of a symbol $\Box\!\!\rightarrow$:

kangaroos have no tails $\Box\!\!\rightarrow$ kangaroos topple over

Counterfactuals—Motivation

Lewis opens his book on counterfactuals with the following statement:

> '*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. I shall give a general analysis of counterfactual conditionals along these lines. (Lewis, 1973)

Counterfactual Conditionals and Indicative Conditionals

- Philosophers make a distinction between indicative conditionals and counterfactuals. Compare:
  (a) If Oswald did not kill Kennedy, then someone else did.
  (b) If Oswald had not killed Kennedy, then someone else would have.
- The indicative/counterfactual distinction is based on the grammar of Latin; modern grammarians talk about open and remote modalities.
- Lewis 1973 points out a difference in truth conditions of (a) and (b) and says his theory applies only to counterfactuals. But Stalnaker (1968) gives a very similar theory for conditionals in general.

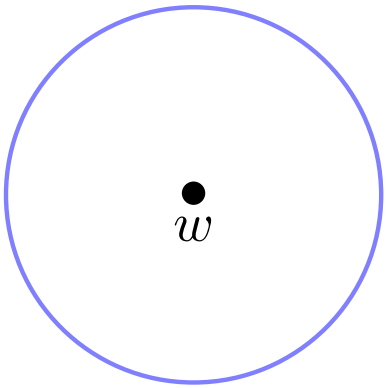
Is the Counterfactual a Strict Conditional?

- Counterfactuals obviously are not material implications—a sentence like *If John had pulled the lever the Earth would have stopped turning* may well be false even if John did not pull the lever.
- One idea, pursued by Clarence Irving Lewis (1883-1964), is that conditional sentences *if* φ *then* ψ have the form $\varphi \mathrel{-\!\!\!\supset} \psi$, where this is an abbreviation of $\Box(\varphi \rightarrow \psi)$ or $[\mathsf{N}](\varphi \rightarrow \psi)$ (historically $\mathrel{-\!\!\!\supset}$ was introduced before $\Box$-operators where).
- Lewis (1973) argues against this approach on the basis of the argument on the following slide.

Why Counterfactuals are not Strict Conditionals

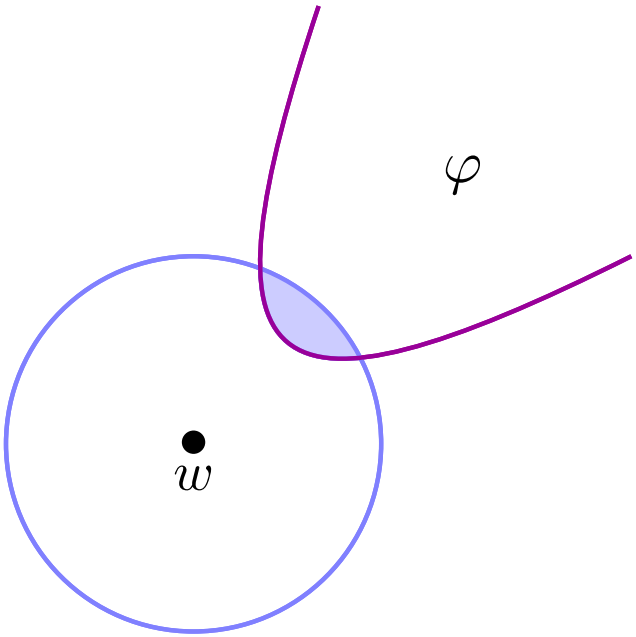
- Consider: *If Otto had come, it would have been a lively party; but if both Otto and Anna had come it would have been a dreary party; but if Waldo had come as well, it would have been lively; but ...*
- Suppose counterfactuals are strict conditionals. Then we can formalise the first two sentences as $[\mathsf{N}](\varphi_1 \to \psi)$ and $[\mathsf{N}]((\varphi_1 \wedge \varphi_2) \to \neg\psi)$ respectively.
- But these two formulas entail $\neg\langle\mathsf{N}\rangle(\varphi_1 \wedge \varphi_2)$, i.e. it is impossible for both Otto and Anna to both come to the party! ↯
- Easy exercise: show $[\mathsf{N}](\varphi_1 \to \psi),\ [\mathsf{N}]((\varphi_1 \wedge \varphi_2) \to \neg\psi) \models \neg\langle\mathsf{N}\rangle(\varphi_1 \wedge \varphi_2)$.
- We also have strengthening the antecedent for strict conditionals: $[\mathsf{N}](\varphi_1 \to \psi) \models [\mathsf{N}]((\varphi_1 \wedge \varphi_2) \to \psi)$, but this should not be valid for counterfactuals.

The Strict Conditional



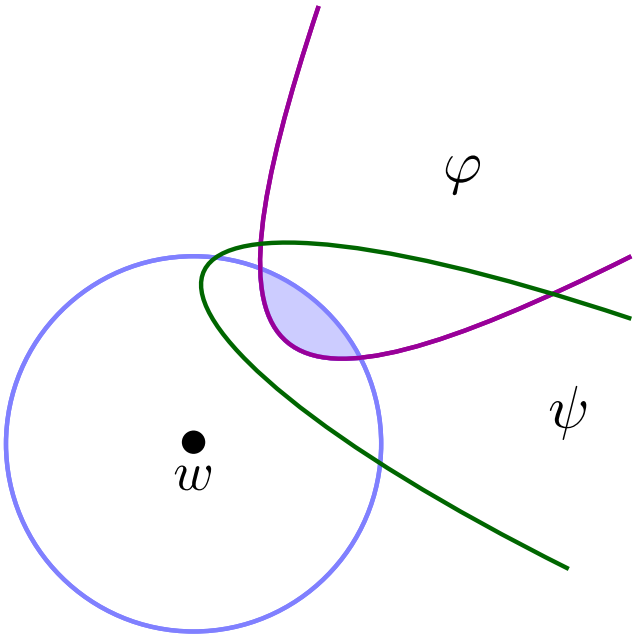
- Consider the world of assessment w and all the worlds N-accessible to it, which form w's *sphere of accessibility*. If the accessibility relation N is reflexive, w is an element of this sphere.

The Strict Conditional



- Consider the world of assessment w and all the worlds N-accessible to it, which form w's *sphere of accessibility*. If the accessibility relation N is reflexive, w is an element of this sphere.
- Consider φ and the accessible φ-worlds.

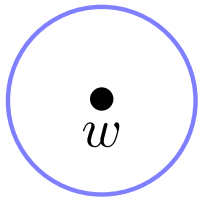
The Strict Conditional



- Consider the world of assessment w and all the worlds N-accessible to it, which form w's *sphere of accessibility*. If the accessibility relation N is reflexive, w is an element of this sphere.
- Consider φ and the accessible φ-worlds.
- If all accessible φ-worlds are ψ-worlds, then $[\mathsf{N}](\varphi \rightarrow \psi)$ holds.

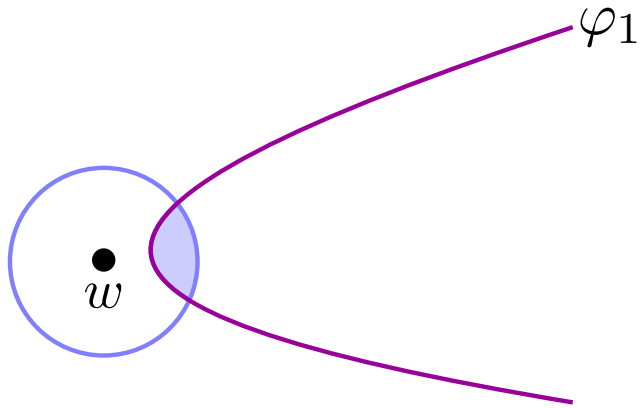
‘Changing Our Minds’ about the Accessibility Relation

Lewis (1973): ‘If we treat the counterfactual as a strict conditional based on similarity, then the best we can do for our troublesome sequences is to keep changing our minds about which such strict conditional it is. We may be able to make two sentences at any one stage true by an appropriate choice of a sphere of accessibility based on similarity, but we must choose anew for each stage.’



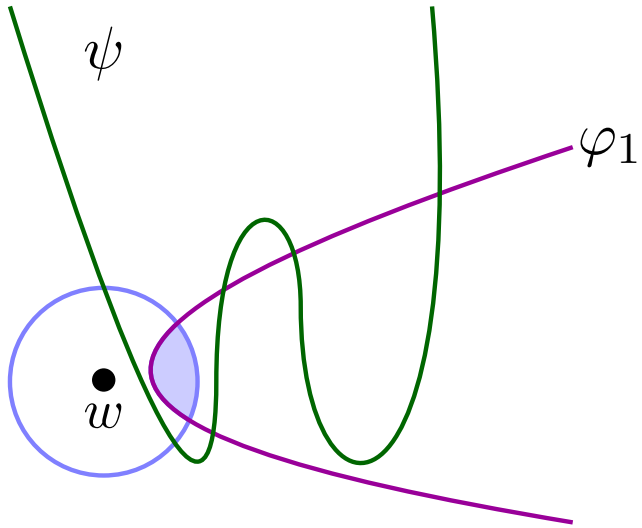
'Changing Our Minds' about the Accessibility Relation

Lewis (1973): 'If we treat the counterfactual as a strict conditional based on similarity, then the best we can do for our troublesome sequences is to keep changing our minds about which such strict conditional it is. We may be able to make two sentences at any one stage true by an appropriate choice of a sphere of accessibility based on similarity, but we must choose anew for each stage.'



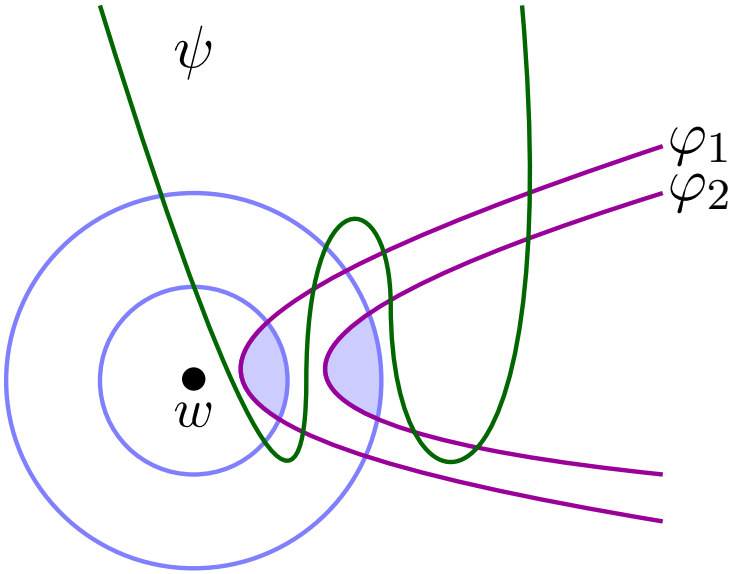
'Changing Our Minds' about the Accessibility Relation

Lewis (1973): 'If we treat the counterfactual as a strict conditional based on similarity, then the best we can do for our troublesome sequences is to keep changing our minds about which such strict conditional it is. We may be able to make two sentences at any one stage true by an appropriate choice of a sphere of accessibility based on similarity, but we must choose anew for each stage.'



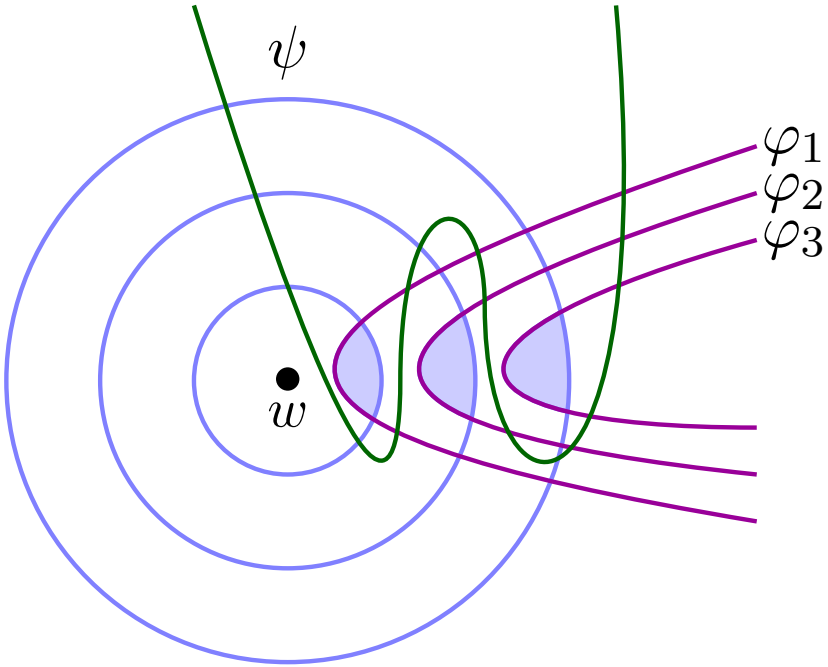
'Changing Our Minds' about the Accessibility Relation

Lewis (1973): 'If we treat the counterfactual as a strict conditional based on similarity, then the best we can do for our troublesome sequences is to keep changing our minds about which such strict conditional it is. We may be able to make two sentences at any one stage true by an appropriate choice of a sphere of accessibility based on similarity, but we must choose anew for each stage.'



‘Changing Our Minds’ about the Accessibility Relation

Lewis (1973): ‘If we treat the counterfactual as a strict conditional based on similarity, then the best we can do for our troublesome sequences is to keep changing our minds about which such strict conditional it is. We may be able to make two sentences at any one stage true by an appropriate choice of a sphere of accessibility based on similarity, but we must choose anew for each stage.’



The Fallacy of Transitivity

That the inference pattern $\varphi \mathrel{\Box\!\!\rightarrow} \psi,\ \psi \mathrel{\Box\!\!\rightarrow} \chi \models \varphi \mathrel{\Box\!\!\rightarrow} \chi$ should not come out valid is shown by the following counterexample from Stalnaker (1968):

- If J. Edgar Hoover had been born a Russian, he would have been a Communist.
- If he had been a Communist, he would have been a traitor.
- Therefore: If he had been born a Russian, he would have been a traitor.

It can easily be checked that $\varphi \mathrel{\neg\!\!\exists} \psi,\ \psi \mathrel{\neg\!\!\exists} \chi \models \varphi \mathrel{\neg\!\!\exists} \chi$ is correct if $\varphi \mathrel{\neg\!\!\exists} \psi$ is short for $[\mathsf{R}](\varphi \rightarrow \psi)$ for some R. So this is another argument that $\mathrel{\Box\!\!\rightarrow}$ cannot be analysed as $\mathrel{\neg\!\!\exists}$.

The Fallacy of Contraposition

$\varphi \boxrightarrow \psi \models \neg\psi \boxrightarrow \neg\chi$ should also be incorrect. Lewis (1973) gives the following counterexample:

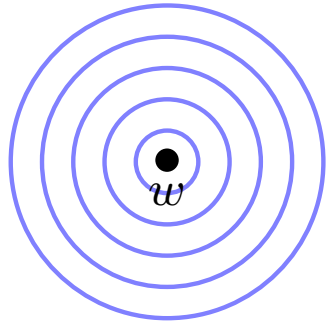
- If Boris had gone to the party, Olga would still have gone.
- Therefore: If Olga had not gone, Boris would still not have gone.

Imagine a situation in which Olga is eager to meet Boris, while Boris wanted to go but stayed away to avoid Olga.

Since $\varphi \mathrel{-\!\!\!\ni} \psi \models \neg\psi \mathrel{-\!\!\!\ni} \neg\chi$, this is another argument against treating $\boxrightarrow$ as $\mathrel{-\!\!\!\ni}$.

Lewis's Systems of Spheres

Lewis considers systems of spheres $ that assign to each possible world a set of sets of possible worlds $\$_w$ with the following properties:



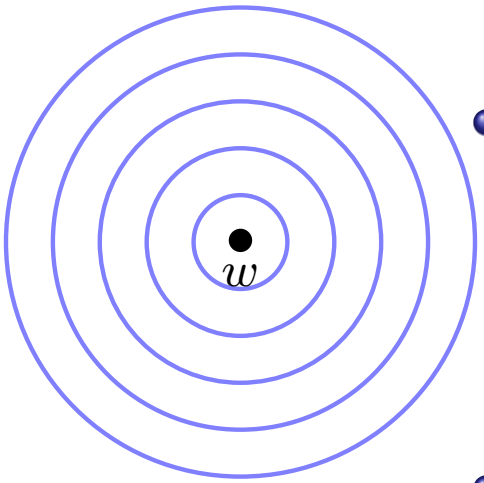
(C) $\$_w$ is centered on w; that is the set $\{w\}$ having w as its only member belongs to $\$_w$;

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

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

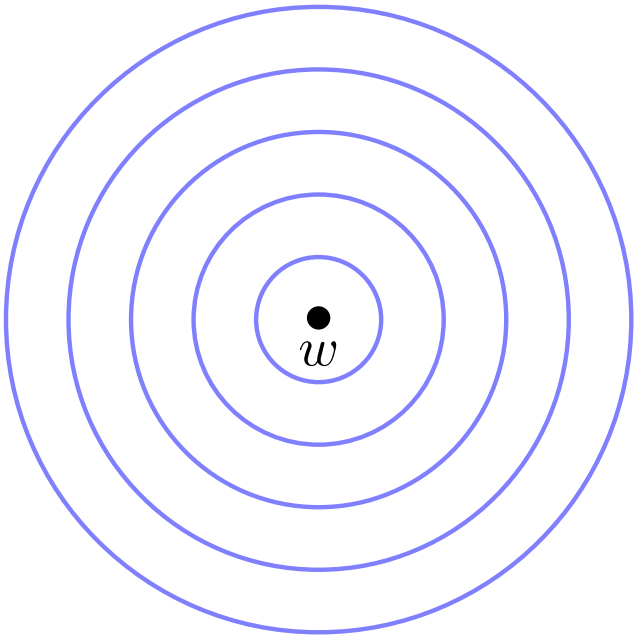
(3) $\$_w$ is closed under (nonempty) intersections; that is, whenever $\mathcal{S}$ is a nonempty subset of $\$_w$ and $\bigcap\mathcal{S}$ is the set of all worlds w' such that w' belongs to all members of $\mathcal{S}$, $\bigcap\mathcal{S}$ belongs to $\$_w$;

Lewis's Systems of Spheres—Truth Definition



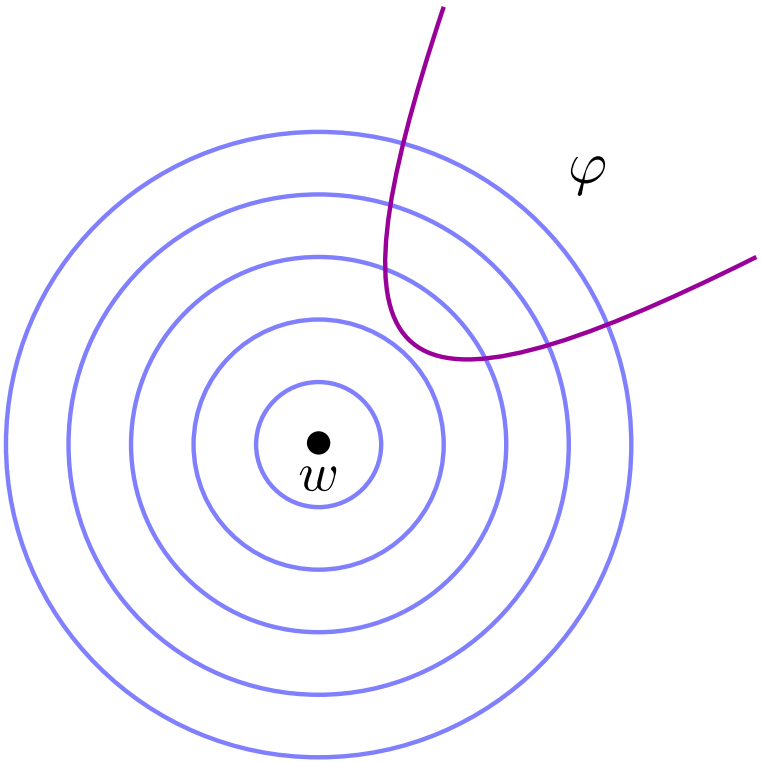
- Whenever one world lies within some sphere around w and another world lies outside that sphere, the first world is more closely similar to w than the second.
- The idea is now that $\varphi \boxdot\!\!\rightarrow \psi$ is true at w iff either
  1. no φ-world belongs to any sphere S in $\$_w$, or
  2. some sphere S in $\$_w$ does contain at least one φ-world and $\varphi \rightarrow \psi$ holds at every world in S.
- The second condition roughly says that the closest φ-worlds are also ψ-worlds.
- We will flesh out the consequences of this definition in the next slides.

Lewis's Systems of Spheres—Non-vacuous Truth



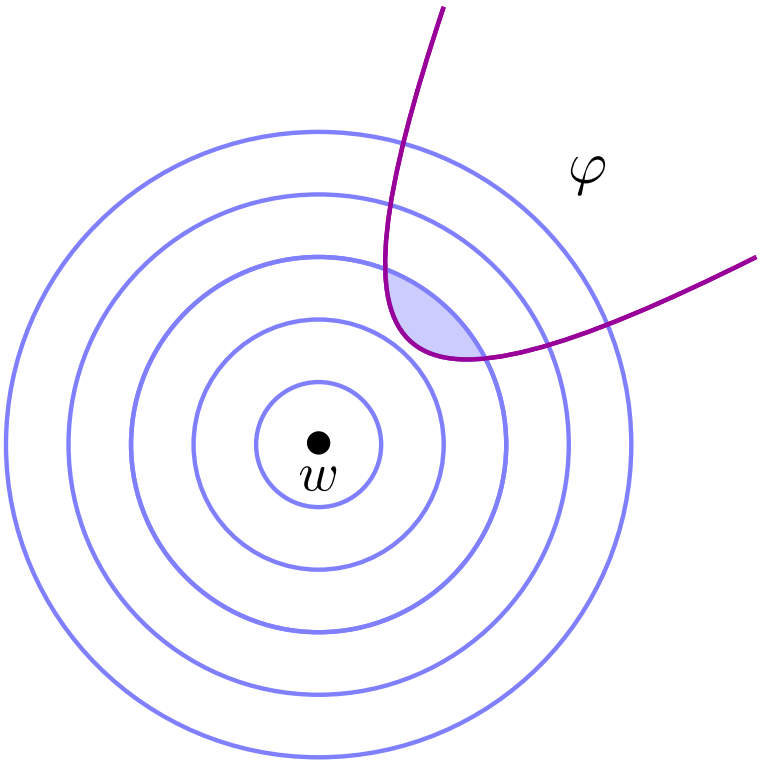
- The set of spheres around w;

Lewis's Systems of Spheres—Non-vacuous Truth



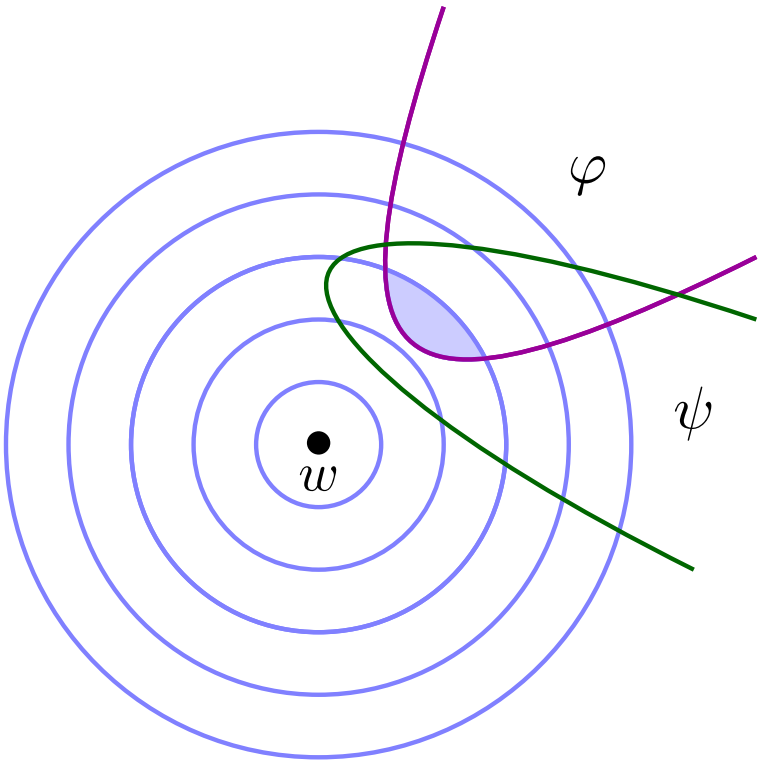
- The set of spheres around w;
- Consider φ;

Lewis's Systems of Spheres—Non-vacuous Truth



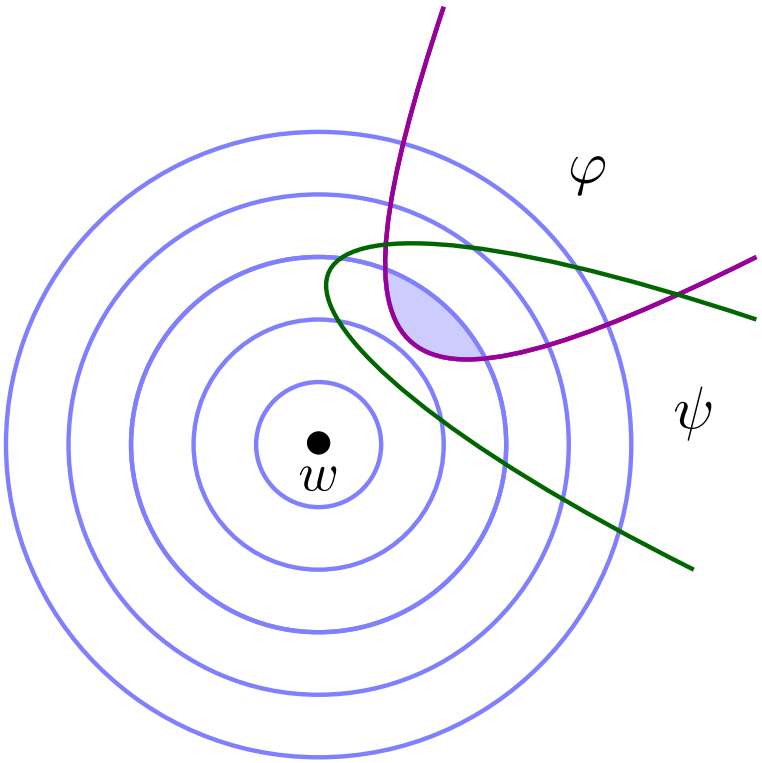
- The set of spheres around w;
- Consider φ;
- In this case the third sphere, let's call it S_3, contains φ-worlds;

Lewis's Systems of Spheres—Non-vacuous Truth



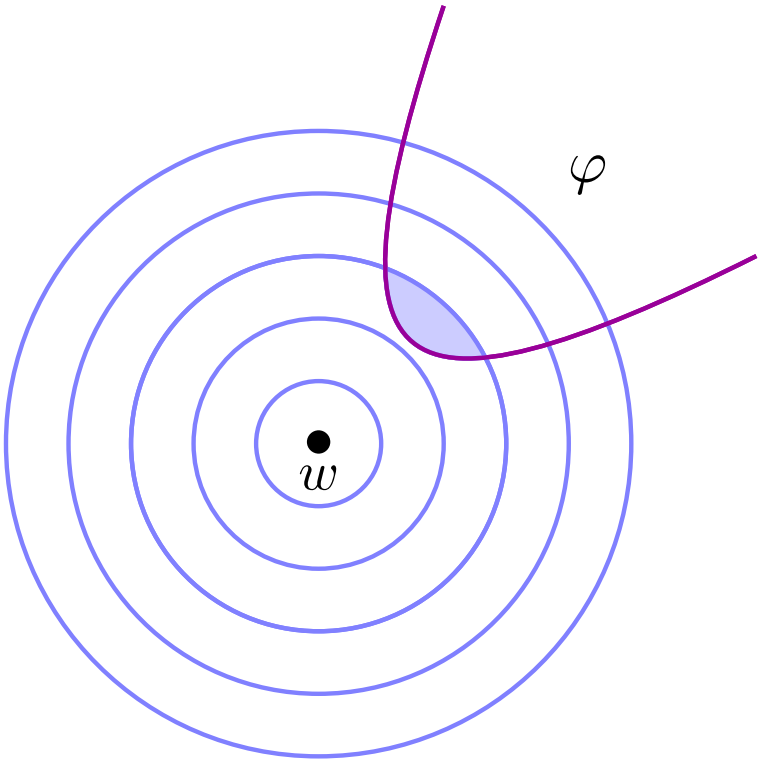
- The set of spheres around w;
- Consider φ;
- In this case the third sphere, let's call it S_3, contains φ-worlds;
- If all these φ-worlds are also ψ-worlds, then $\varphi \to \psi$ holds at every world in S_3, so that $\varphi \boxdot\!\!\rightarrow \psi$ is true;

Lewis's Systems of Spheres—Non-vacuous Truth



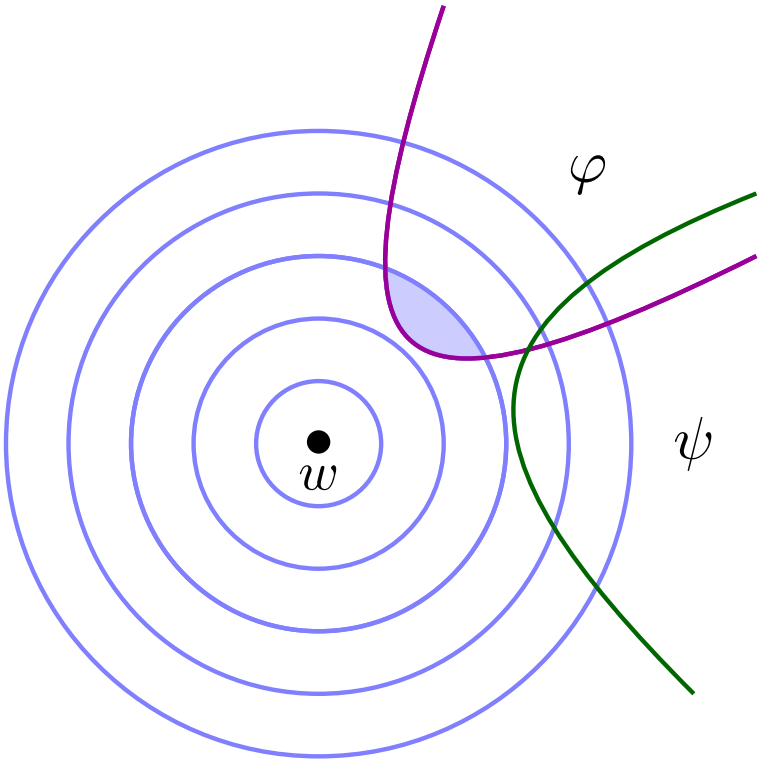
- The set of spheres around w;
- Consider φ;
- In this case the third sphere, let's call it S_3, contains φ-worlds;
- If all these φ-worlds are also ψ-worlds, then $\varphi \to \psi$ holds at every world in S_3, so that $\varphi \boxdot\!\!\!\rightarrow \psi$ is true;
- We also have $\neg(\varphi \boxdot\!\!\!\rightarrow \neg\psi)$ in this case.

Lewis's Systems of Spheres—Falsity, Opposite True



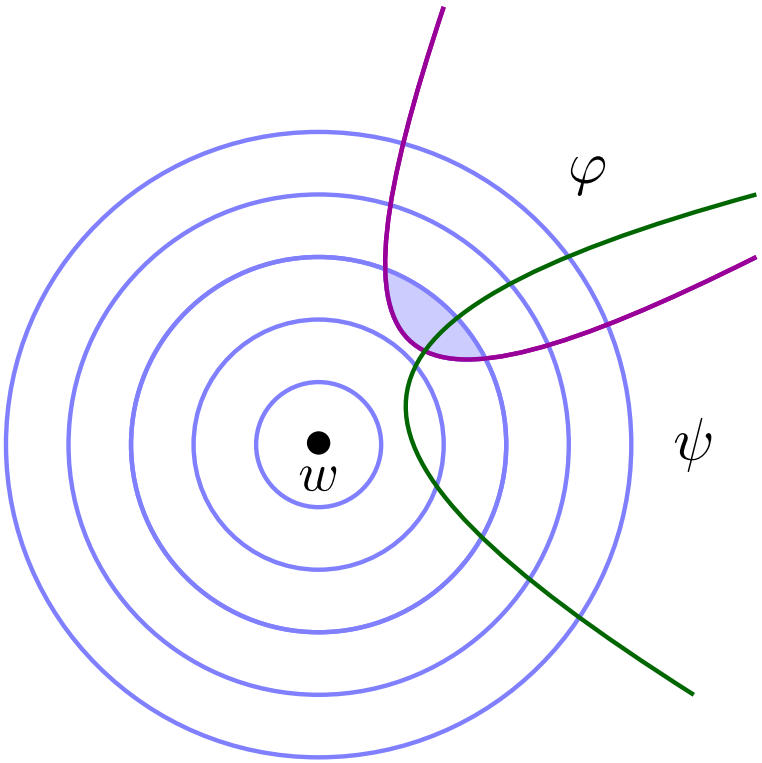
- The system of spheres, φ, and the φ-worlds in S_3 again;

Lewis's Systems of Spheres—Falsity, Opposite True



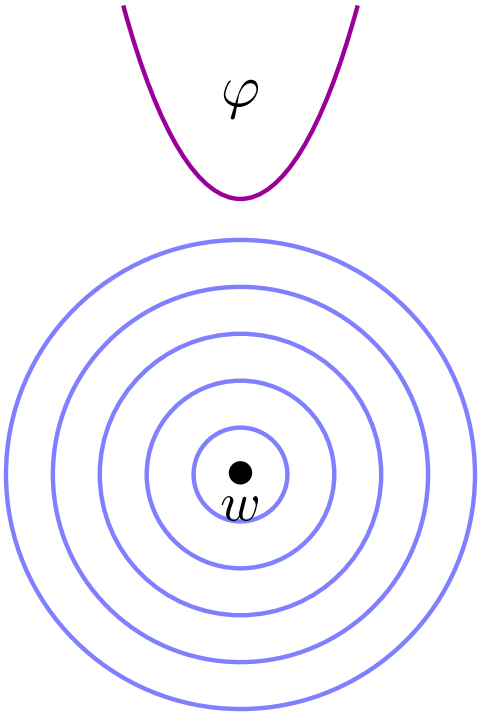
- The system of spheres, φ, and the φ-worlds in S_3 again;
- This time, while there is a sphere (S_3) containing φ-worlds, there is no such sphere such that $\varphi \rightarrow \psi$ holds in it;
- We therefore have $\neg(\varphi \boxdot\!\!\rightarrow \psi)$;
- $\varphi \boxdot\!\!\rightarrow \neg\psi$ also holds—all φ-worlds in S_3 are $\neg\psi$-worlds.

Lewis's Systems of Spheres—Falsity, Opposite False



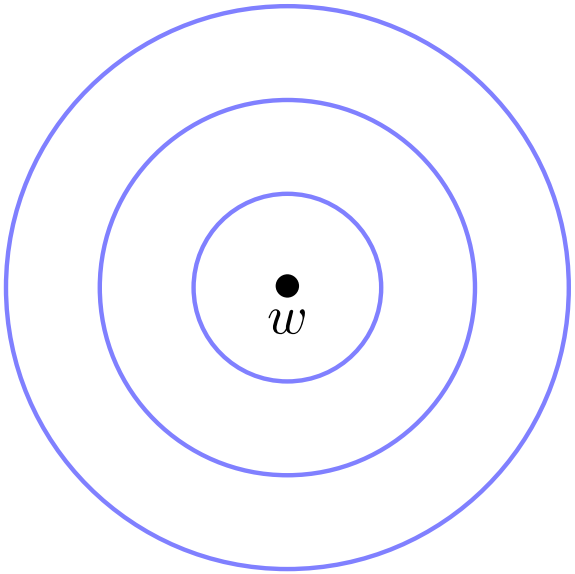
- Let's have a look at the case where some, but not all, of the closest φ-worlds are ψ-worlds;
- Then $\neg(\varphi \boxdot\!\!\rightarrow \psi)$ and $\neg(\varphi \boxdot\!\!\rightarrow \neg\psi)$ are both true.

Lewis's Systems of Spheres—Vacuous Truth

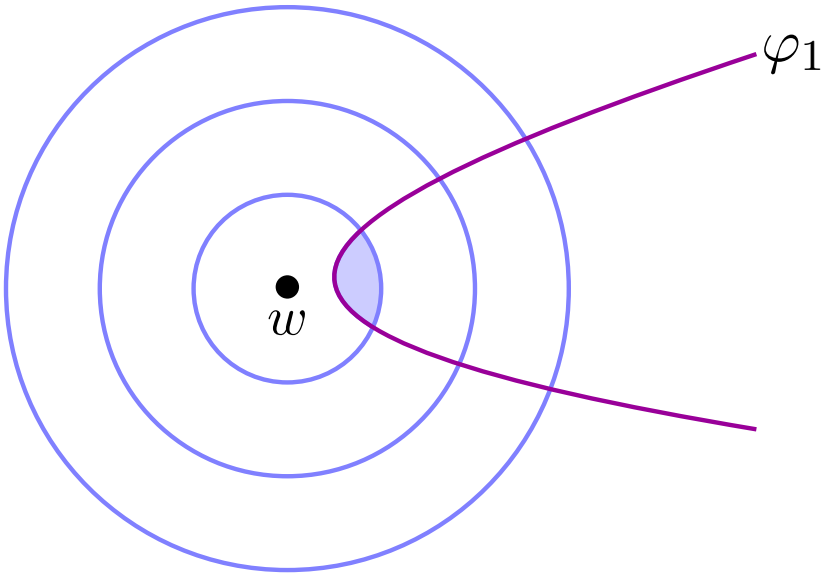


- In the (degenerate) case in which there are no accessible φ-worlds, $\varphi \boxrightarrow \chi$ is true for any χ. So both $\varphi \boxrightarrow \psi$ and $\varphi \boxrightarrow \neg\psi$ come out true.
- Lewis also discusses an alternative operator $\Box\!\!\Rightarrow$ so that $\varphi \Box\!\!\Rightarrow \psi$ is defined to be true at w iff some sphere S in $\$_w$ does contain at least one φ-world and $\varphi \rightarrow \psi$ holds at every world in S.
- $\varphi \Box\!\!\Rightarrow \psi$ and $\varphi \Box\!\!\Rightarrow \neg\psi$ both come out false in the case sketched here.

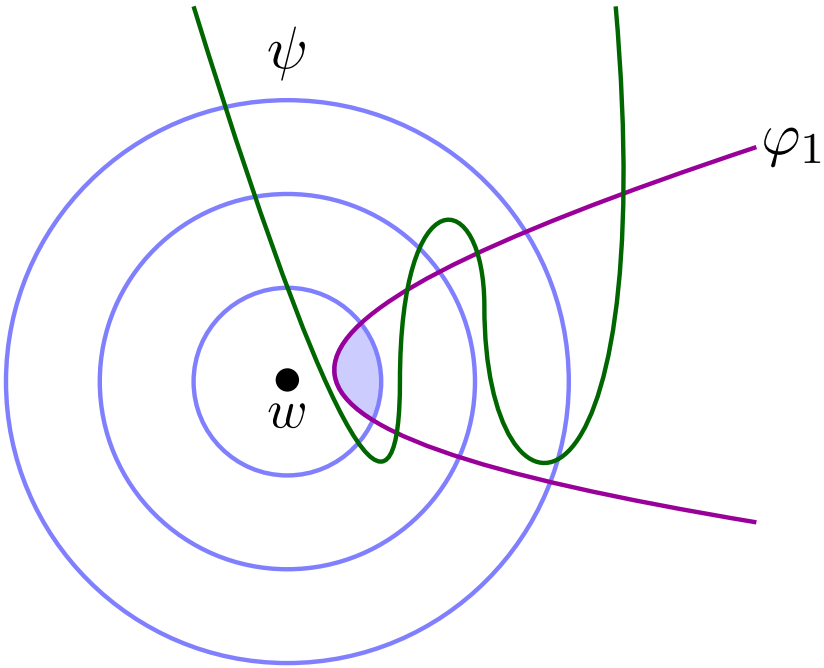
Lewis's Systems of Spheres—Nonmonotonicity



Lewis's Systems of Spheres—Nonmonotonicity



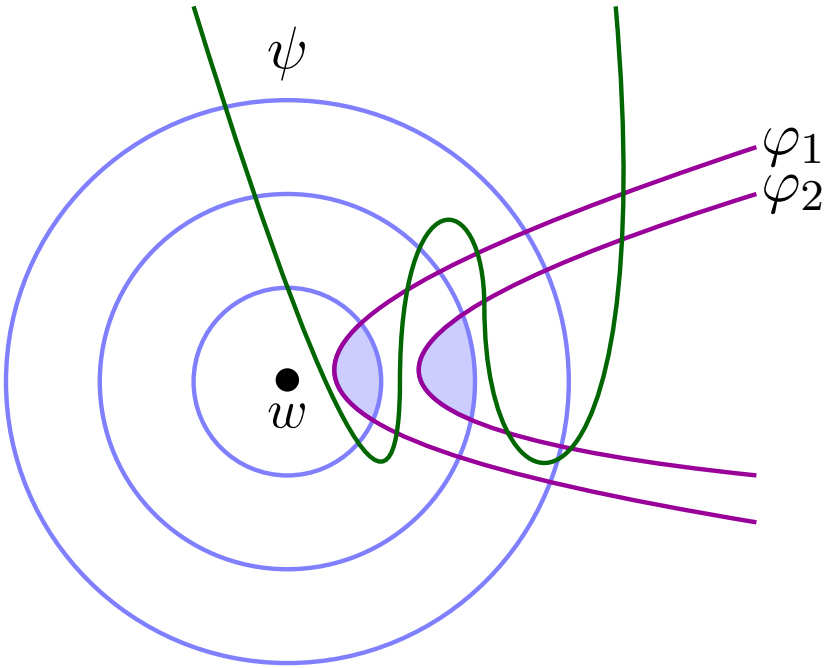
Lewis's Systems of Spheres—Nonmonotonicity



$\varphi_1 \boxdot\!\!\rightarrow \psi$

$\neg(\varphi_1 \boxdot\!\!\rightarrow \neg\psi)$

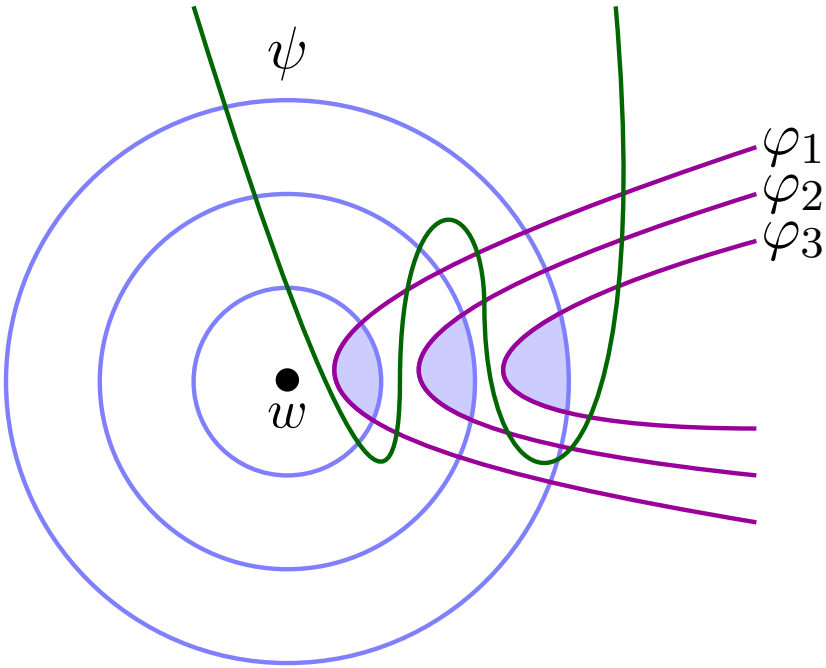
Lewis's Systems of Spheres—Nonmonotonicity



$\varphi_1 \boxdot\!\!\rightarrow \psi$
$\neg(\varphi_1 \boxdot\!\!\rightarrow \neg\psi)$

$(\varphi_1 \wedge \varphi_2) \boxdot\!\!\rightarrow \neg\psi$
$\neg((\varphi_1 \wedge \varphi_2) \boxdot\!\!\rightarrow \psi)$

Lewis's Systems of Spheres—Nonmonotonicity



$\varphi_1 \boxrightarrow \psi$
$\neg(\varphi_1 \boxrightarrow \neg\psi)$

$(\varphi_1 \wedge \varphi_2) \boxrightarrow \neg\psi$
$\neg((\varphi_1 \wedge \varphi_2) \boxrightarrow \psi)$

$(\varphi_1 \wedge \varphi_2 \wedge \varphi_2) \boxrightarrow \psi$
$\neg((\varphi_1 \wedge \varphi_2 \wedge \varphi_2) \boxrightarrow \neg\psi)$

Basing the System on a Comparative Similarity Relation

- Lewis gives several reformulations of his theory. One is based on a comparative similarity relation: $w_1 \leqslant_w w_2$ means that w_1 is as least as similar to w as w_2 is. For our purposes it is better to take the converse of this relation and from now on we will write this as $w_2 \geqslant^s_w w_1$ (the 's' is for similarity).
- $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ is now defined to be true at w iff either
  1. no φ-world is accessible from w, or
  2. there is an accessible φ-world w_1 such that, for any world w_2, if $w_1 \geqslant^s_w w_2$ then $\varphi \rightarrow \psi$ holds at w_2.

Translation into Predicate Logic

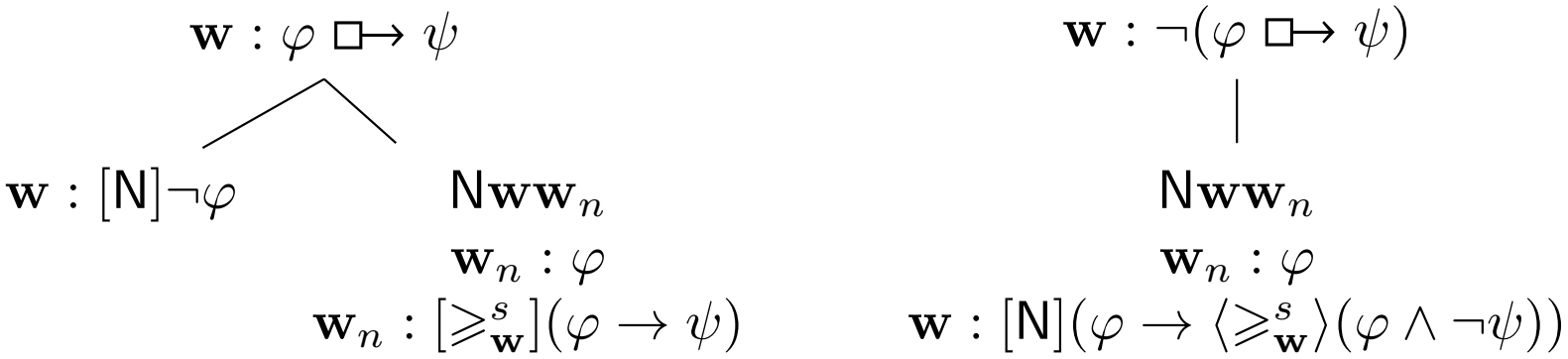
- Up till now we have treated modal operators by considering pairs $w : \varphi$ (where φ is a sentence in our modal language) as shorthand for predicate logical formulas. Lewis's reformulation of his theory in terms of similarity relations enables to extend our definition with a clause for $\varphi \boxdot\!\!\rightarrow \psi$.
- $w : \varphi \boxdot\!\!\rightarrow \psi \triangleq \forall w'(\mathsf{N}ww' \rightarrow w' : \neg\varphi) \vee$
  $\exists w'(\mathsf{N}ww' \wedge w' : \varphi \wedge \forall w''(w' \geqslant^s_w w'' \rightarrow w'' : (\varphi \rightarrow \psi)))$
- If we treat $\geqslant^s_w$ as just another accessibility relation, we have $w : \varphi \boxdot\!\!\rightarrow \psi$ iff $w : [\mathsf{N}]\neg\varphi \vee \langle\mathsf{N}\rangle(\varphi \wedge [\geqslant^s_w](\varphi \rightarrow \psi))$.
- We have chosen N as our accessibility relation, but nothing hinges on this.

Adding Axioms

- In order to ensure that the logic based on N and $\geqslant^s$ does what systems of spheres do, we must impose axioms:
- $\forall w w_1 w_2 w_3((w_1 \geqslant^s_w w_2 \wedge w_2 \geqslant^s_w w_3) \rightarrow w_1 \geqslant^s_w w_3)$
  $\forall w w_1\, w_1 \geqslant^s_w w_1$
  $\forall w w_1 w_2((\mathsf{N} w w_1 \wedge w_1 \geqslant^s_w w_2) \rightarrow \mathsf{N} w w_2)$
  $\forall w_1\, \mathsf{N} w_1 w_1$
  $\forall w w_1(w \geqslant^s_w w_1 \rightarrow w = w_1)$
  $\forall w w_1 w_2(w_1 \geqslant^s_w w_2 \vee w_2 \geqslant^s_w w_1)$
- (It can be shown that from any structure satisfying these axioms a system of spheres can be derived.)
- We will compile these axioms into tableau rules shortly.

Tableau Rules for $\Box\!\!\rightarrow$

Our translation leads to the following tableau rules.



In both cases $\mathbf{w}_n$ must be new.

Tableau Rules for N and $\geqslant^s$

Transitivity $\geqslant^s_{\mathbf{w}}$

$$\mathbf{w}_1 \geqslant^s_{\mathbf{w}} \mathbf{w}_2,\ \mathbf{w}_2 \geqslant^s_{\mathbf{w}} \mathbf{w}_3 \quad \Big| \quad \mathbf{w}_1 \geqslant^s_{\mathbf{w}} \mathbf{w}_3$$

Reflexivity $\geqslant^s_{\mathbf{w}}$

$$\Big| \quad \mathbf{w}_1 \geqslant^s_{\mathbf{w}} \mathbf{w}_1$$

Downward Closure

$$\mathsf{N}\mathbf{w}\mathbf{w}_1,\ \mathbf{w}_1 \geqslant^s_{\mathbf{w}} \mathbf{w}_2 \quad \Big| \quad \mathsf{N}\mathbf{w}\mathbf{w}_2$$

Reflexivity N

$$\Big| \quad \mathsf{N}\mathbf{w}\mathbf{w}$$

Strict Minimality

$$\mathbf{w} \geqslant^s_{\mathbf{w}} \mathbf{w}_1 \quad \Big| \quad \mathbf{w} = \mathbf{w}_1$$

Connectivity $\geqslant^s_{\mathbf{w}}$

$$\mathbf{w}_1 \geqslant^s_{\mathbf{w}} \mathbf{w}_2 \qquad \mathbf{w}_2 \geqslant^s_{\mathbf{w}} \mathbf{w}_1$$

Calculemus!

- We now have a tableau system and we can find out what it predicts.
- Show the following:
  $\models_{AX} \varphi \boxdot\!\!\rightarrow \varphi$
  $\models_{AX} (\neg\varphi \boxdot\!\!\rightarrow \varphi) \rightarrow (\psi \boxdot\!\!\rightarrow \varphi)$
  $\models_{AX} (\varphi \boxdot\!\!\rightarrow \neg\psi) \vee (((\varphi \wedge \psi) \boxdot\!\!\rightarrow \chi) \leftrightarrow (\varphi \boxdot\!\!\rightarrow (\psi \rightarrow \chi)))$
  $\models_{AX} (\varphi \boxdot\!\!\rightarrow \psi) \rightarrow (\varphi \rightarrow \psi)$
  $\models_{AX} (\varphi \wedge \psi) \rightarrow (\varphi \boxdot\!\!\rightarrow \psi)$
- Show failure of Transitivity, Contraposition, and Strengthening the Antecedent:
  $\varphi \boxdot\!\!\rightarrow \psi,\ \psi \boxdot\!\!\rightarrow \chi \not\models_{AX} \varphi \boxdot\!\!\rightarrow \chi$
  $\varphi \boxdot\!\!\rightarrow \psi \not\models_{AX} \neg\psi \boxdot\!\!\rightarrow \neg\varphi$
  $(\varphi_1 \boxdot\!\!\rightarrow \psi) \not\models_{AX} ((\varphi_1 \wedge \varphi_2) \boxdot\!\!\rightarrow \psi)$

Deontic Logic Revisited

- In a previous lecture the system SDL of Standard Deontic Logic was introduced. It consisted of the usual rules for a box operator [O], plus Seriality.
- If furthermore Shift Reflexivity is added, a system SDL^+ is obtained. The following are tableau rules for Seriality and Shift Reflexivity.

($\mathbf{w}$ old $\mathbf{w}_n$ new)

- Are these rules adequate at all? Paul McNamara's item in the *Stanford Encyclopedia* gives many problematic cases. In the next slides we will concentrate on Chisholm's Contrary-to-Duty Paradox and Forrester's Paradox of the Gentle Murderer.

Chisholm's Paradox

Consider the following scenario (Chisholm 1963).

1. It ought to be that Jones goes to the assistance of his neighbours.
2. It ought to be that if Jones goes, then he tells them he is coming.
3. If Jones doesn't go, then he ought not tell them he is coming.
4. Jones doesn't go.

These four sentences seem to be mutually consistent and logically independent (none of them follows from the rest). Can we find a formalisation with the help of $[\mathsf{O}]$ that retains these features?

Chisholm's Paradox—Straightforward Formalisation

1. It ought to be that Jones goes to the assistance of his neighbours. $[\mathsf{O}]G$
2. It ought to be that if Jones goes, then he tells them he is coming. $[\mathsf{O}](G \to T)$
3. If Jones doesn't go, then he ought not tell them he is coming. $\neg G \to [\mathsf{O}]\neg T$
4. Jones doesn't go. $\neg G$

Unfortunately $\{[\mathsf{O}]G,\ [\mathsf{O}](G \to T),\ \neg G \to [\mathsf{O}]\neg T,\ \neg G\}$ is inconsistent (Exercise: show this using Seriality), so this cannot be an adequate formalisation.

Chisholm's Paradox—Second Try

Perhaps the third premise should be formalised with a wide scope [O], just like the second premise?

1. It ought to be that Jones goes to the assistance of his neighbours. $[\mathsf{O}]G$
2. It ought to be that if Jones goes, then he tells them he is coming. $[\mathsf{O}](G \to T)$
3. If Jones doesn't go, then he ought not tell them he is coming. $[\mathsf{O}](\neg G \to \neg T)$
4. Jones doesn't go. $\neg G$

Now the problem is that 3 follows from 1, as can easily be shown. Again, no adequate formalisation...

Chisholm's Paradox—Third Attempt

Then maybe the second premise should be rendered along the lines of the third?

1. It ought to be that Jones goes to the assistance of his neighbours. $[\mathsf{O}]G$
2. It ought to be that if Jones goes, then he tells them he is coming. $G \to [\mathsf{O}]T$
3. If Jones doesn't go, then he ought not tell them he is coming. $\neg G \to [\mathsf{O}]\neg T$
4. Jones doesn't go. $\neg G$

Now 2 follows from 4 by propositional logic. Again, no independence and no adequate formalisation.

The Gentle Murderer

The Paradox of the Gentle Murderer (Forrester 1984) gives another scenario that seems impossible to formalise with the help of Standard Deontic Logic.

1. It is obligatory that John Doe does not kill his mother. $[\mathsf{O}]\neg K$
2. If Doe does kill his mother, then it is obligatory that Doe kills her gently. $K \to [\mathsf{O}]G$
3. Doe does kill his mother. K
4. (Gentle killing is killing) $[\mathsf{A}](G \to K)$

$\{[\mathsf{O}]\neg K,\ K \to [\mathsf{O}]G,\ K,\ [\mathsf{A}](G \to K)\}$ is inconsistent (check), but the original scenario does not seem to be...

Conditional Obligation

- The jury is still very much out on the puzzles just sketched.
- One set of researchers is defending the position that the material implication $\rightarrow$ in some of the previous formalisations should be replaced by an operator along the lines of Lewis's $\Box\!\!\rightarrow$ (or Stalnaker's operator, which is similar).
- Others have taken the view that a sentence like *if Doe kills his mother, he ought to kill her gently* has a form $\mathsf{OB}(G \mid K)$, where the two-place OB is an operator in its own right.
- Actually, in Lewis's book on Counterfactuals (Lewis 1973), such an operator is also considered.

Strengthening the Antecedent

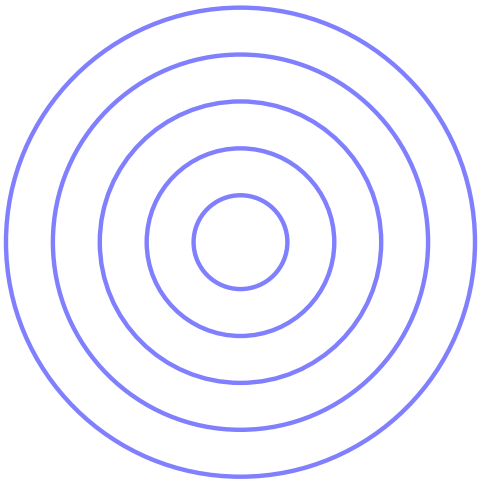
Lewis 1973 gives the following example.

- Given that Jesse robbed the bank, he ought to confess;
- but given in addition that his confession would send his ailing mother to an early grave, he ought not to;
- but given in addition that an innocent man is on trial for the crime, he ought to after all...

We see that Strengthening the Antecedent fails for conditional obligation.

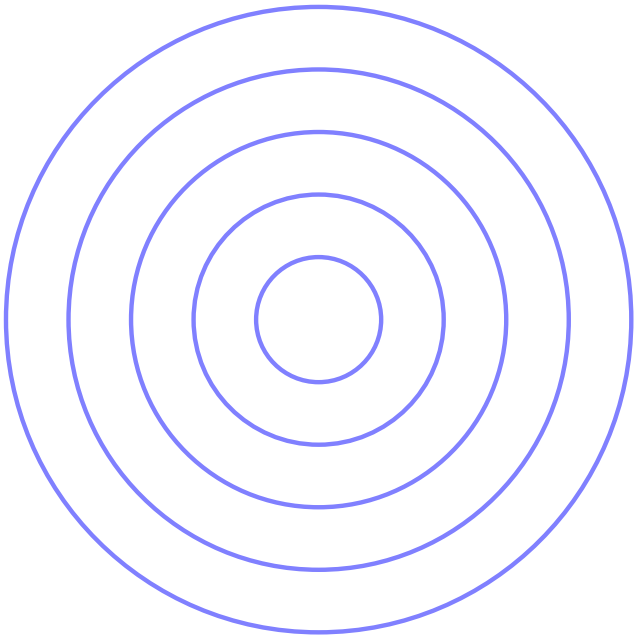
(A question now posits itself: Do Transitivity and Contraposition also fail?)

Conditional Obligation in a System of Spheres—Truth Definition



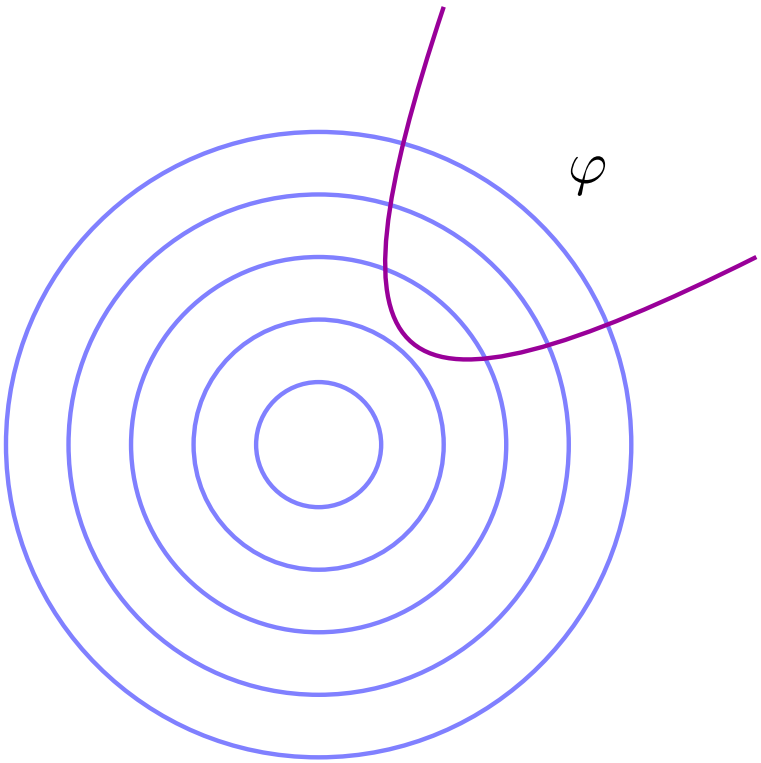
- Systems of spheres are as before, but are based on a betterness relation, not similarity. The actual world is no longer in the centre.
- The idea is now that $\mathsf{OB}(\psi \mid \varphi)$ is true at w iff either
  1. no φ-world belongs to any sphere S in $\$_w$, or
  2. some sphere S in $\$_w$ does contain at least one φ-world and $\varphi \rightarrow \psi$ holds at every world in S.
- The second condition roughly says that the best φ-worlds are also ψ-worlds.

Conditional Obligation in a System of Spheres



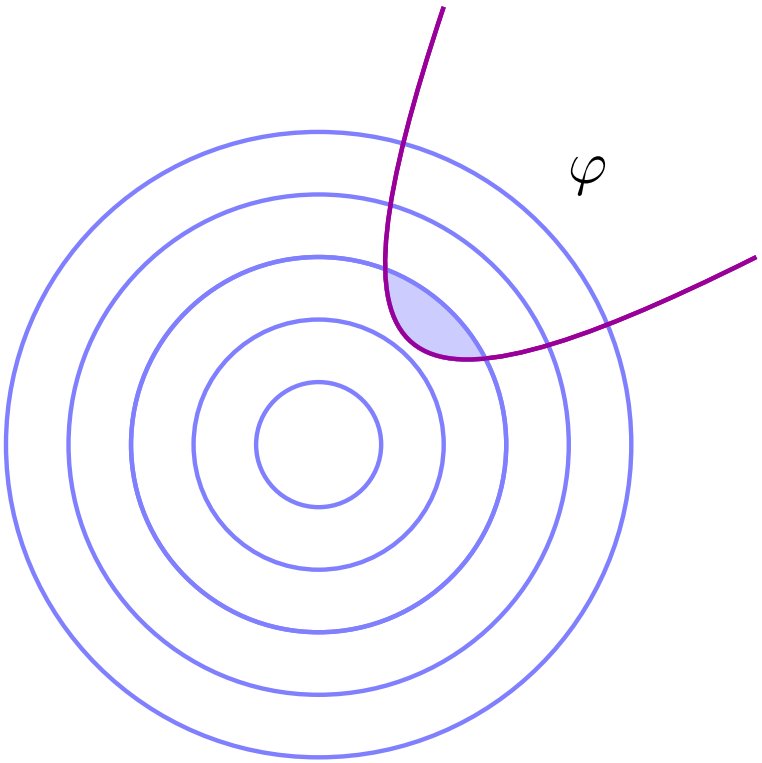
- The set of spheres 'around' w;

Conditional Obligation in a System of Spheres



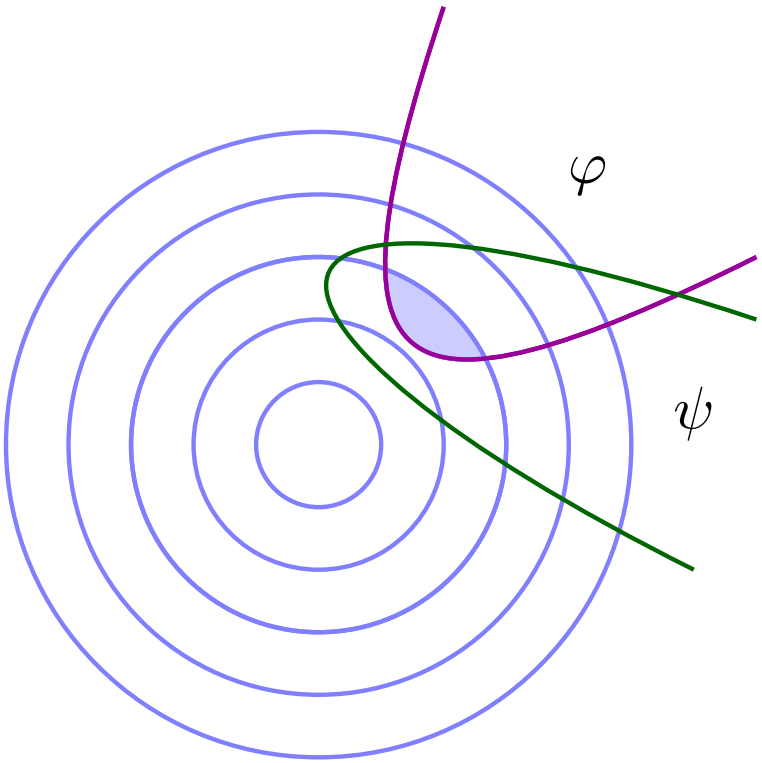
- The set of spheres 'around' w;
- Consider φ;

Conditional Obligation in a System of Spheres



- The set of spheres 'around' w;
- Consider φ;
- In this case the third sphere, let's call it S_3, contains φ-worlds;

Conditional Obligation in a System of Spheres



- The set of spheres 'around' w;
- Consider φ;
- In this case the third sphere, let's call it S_3, contains φ-worlds;
- If all these φ-worlds are also ψ-worlds, then $\varphi \to \psi$ holds at every world in S_3, so that $\mathsf{OB}(\psi \mid \varphi)$ is true;

Basing the System on a Comparative Goodness Relation

- Let's have what Lewis calls a 'comparative goodness' relation $\geqslant^g_w$ (the 'g' is for goodness). The idea is that $w_2 \geqslant^g_w w_1$ iff, from the viewpoint of w, world w_1 is at least as good as w_2.
- $\mathsf{OB}(\psi \mid \varphi)$ is now defined to be true at w iff either
  1. no φ-world is accessible from w, or
  2. there is an accessible φ-world w_1 such that, for any world w_2, if $w_1 \geqslant^g_w w_2$ then $\varphi \rightarrow \psi$ holds at w_2.

Translation into Predicate Logic

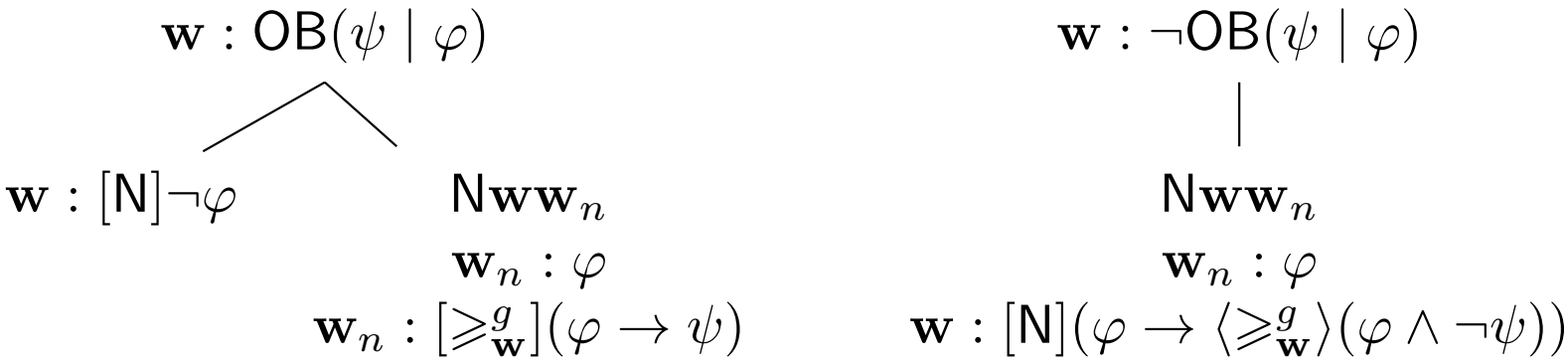
- The following clause translates pairs $w : \mathsf{OB}(\psi \mid \varphi)$ into predicate logic.
- $w : \mathsf{OB}(\psi \mid \varphi) \triangleq \forall w'(\mathsf{N}ww' \to w' : \neg\varphi) \vee$
  $\exists w'(\mathsf{N}ww' \wedge w' : \varphi \wedge \forall w''(w' \geqslant_w^g w'' \to w'' : (\varphi \to \psi)))$
- We have $w : \mathsf{OB}(\psi \mid \varphi)$ iff $w : [\mathsf{N}]\neg\varphi \vee \langle\mathsf{N}\rangle(\varphi \wedge [\geqslant_w^g](\varphi \to \psi))$.
- The treatment thus far is completely analogous to that of $\Box\!\!\rightarrow$.

Unconditional Obligation

- We can define unconditional obligation in terms of conditional obligation. Let $\top$ be some arbitrary tautology (e.g. $P \vee \neg P$). Then $\mathsf{OB}(\varphi)$ is short for $\mathsf{OB}(\varphi \mid \top)$.
- We have $w : \mathsf{OB}(\varphi)$ iff $w : [\mathsf{N}]\neg\top \vee \langle \mathsf{N} \rangle [\geqslant^g_w](\varphi)$. If N is reflexive this boils down to $\langle \mathsf{N} \rangle [\geqslant^g_w](\varphi)$.

Tableau Rules for OB

Our translation leads to the following tableau rules.



In both cases $\mathbf{w}_n$ must be new.

Adding Axioms

- In order to ensure that the logic based on N and $\geqslant^g$ does what it must do, we must impose axioms:
- $\forall w w_1 w_2 w_3((w_1 \geqslant^g_w w_2 \wedge w_2 \geqslant^g_w w_3) \rightarrow w_1 \geqslant^g_w w_3)$
  $\forall w w_1\,(w_1 \geqslant^g_w w_1$
  $\forall w w_1 w_2((\mathsf{N} w w_1 \wedge w_1 \geqslant^g_w w_2) \rightarrow \mathsf{N} w w_2)$
  $\forall w_1\,\mathsf{N} w_1 w_1$
  $\forall w w_1 w_2(w_1 \geqslant^g_w w_2 \vee w_2 \geqslant^g_w w_1)$
- There is no analogue of the strict minimality constraint $\forall w w_1(w \geqslant^s_w w_1 \rightarrow w = w_1)$ on $\geqslant^s$.
- Tableau rules corresponding to these axioms follow.

Tableau Rules for N and $\geqslant^g$

Transitivity $\geqslant^g_{\mathbf{w}}$

$$\mathbf{w}_1 \geqslant^g_{\mathbf{w}} \mathbf{w}_2 \quad \mathbf{w}_2 \geqslant^g_{\mathbf{w}} \mathbf{w}_3$$
$$\Big|$$
$$\mathbf{w}_1 \geqslant^g_{\mathbf{w}} \mathbf{w}_3$$

Reflexivity $\geqslant^g_{\mathbf{w}}$

$$\Big|$$
$$\mathbf{w}_1 \geqslant^g_{\mathbf{w}} \mathbf{w}_1$$

Downward Closure

$$\mathsf{N}\mathbf{w}\mathbf{w}_1$$
$$\mathbf{w}_1 \geqslant^g_{\mathbf{w}} \mathbf{w}_2$$
$$\Big|$$
$$\mathsf{N}\mathbf{w}\mathbf{w}_2$$

Reflexivity N

$$\Big|$$
$$\mathsf{N}\mathbf{w}_1\mathbf{w}_1$$

Connectivity $\geqslant^g_{\mathbf{w}}$

$$\mathbf{w}_1 \geqslant^g_{\mathbf{w}} \mathbf{w}_2 \qquad \mathbf{w}_2 \geqslant^g_{\mathbf{w}} \mathbf{w}_1$$

But What About Detachment?

- Note that, given the rules for $\Box\!\!\rightarrow$, the following version of Modus Ponens (aka Detachment) is valid.
- $\varphi \Box\!\!\rightarrow \psi,\ \varphi \models_{AX} \psi$
- Test the following:
  Factual Detachment: $\mathsf{OB}(\psi \mid \varphi),\ \varphi \models_{AX} \mathsf{OB}\psi$
  Deontic Detachment: $\mathsf{OB}(\psi \mid \varphi),\ \mathsf{OB}\varphi \models_{AX} \mathsf{OB}\psi$
- Should the results be as they are predicted here?

Conditional Belief

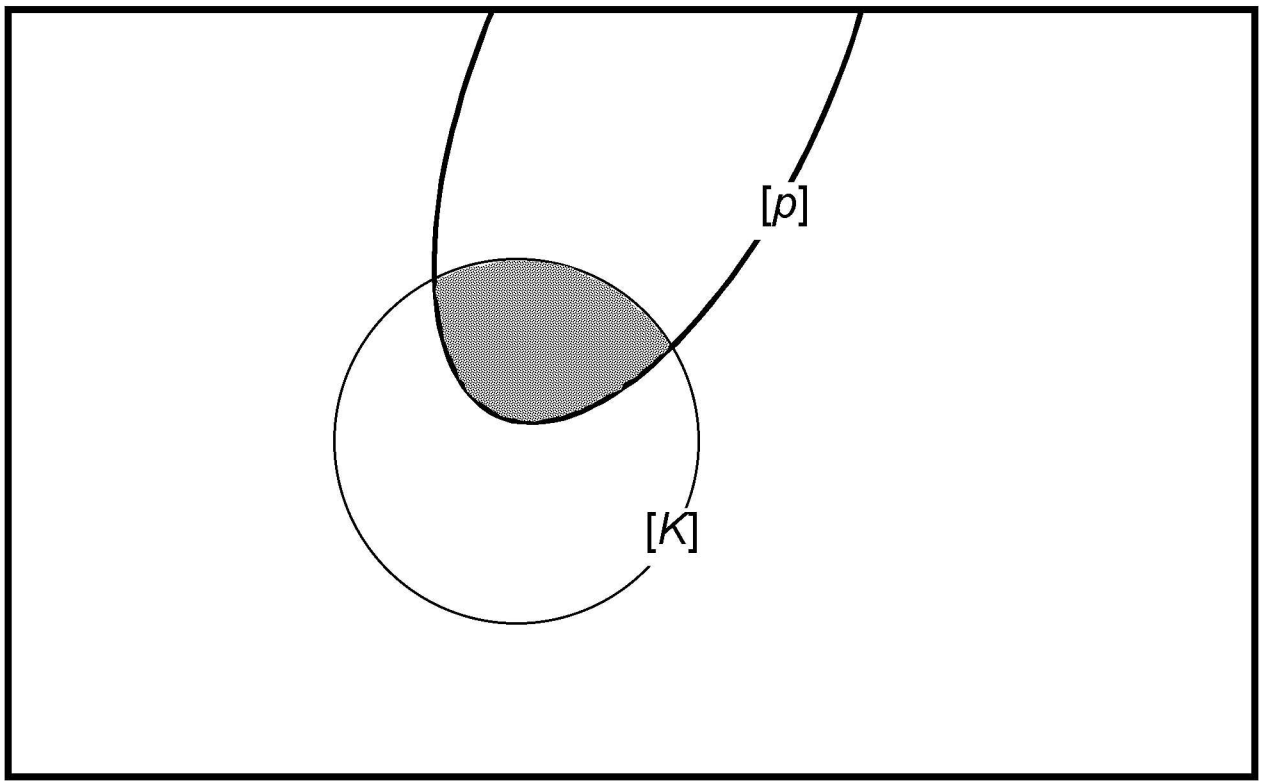
- The Lewis-Stalnaker treatment of counterfactuals and obligation has a general structure that can be applied to many other cases.
- One of those other applications is that of 'soft' belief.
- There may be many things we believe in, but usually some beliefs are deeper entrenched than others. Think of the difference between I believe that the earth is round and I believe that Mary is suffering from the effects of a bad hair day.
- We may also have conditional beliefs like if the meter is in the red area the pressure is too high.
- One way to model conditional beliefs is using probability theory (which can be formalised using logic), but a formalisation along the lines of Lewis's theory is also possible.

Belief Revision

- Philosophers of science and others have traditionally been interested in theory change. How should theories be altered in the light of new evidence? More in general, how should beliefs be changed if new facts come to light?
- Since the early 1980s researchers in the Belief Revision paradigm have been working on theories describing this process.
- The original theories worked with postulates, but in Grove 1988 it was pointed out that in many cases there are equivalent treatments in terms of possible worlds that follow Lewis's ideas.
- The following pictures are illustrative. They were taken from Sven Ove Hansson's item on the Logic of Belief Revision in the Stanford Encyclopedia (`http://plato.stanford.edu/entries/logic-belief-revision/`).

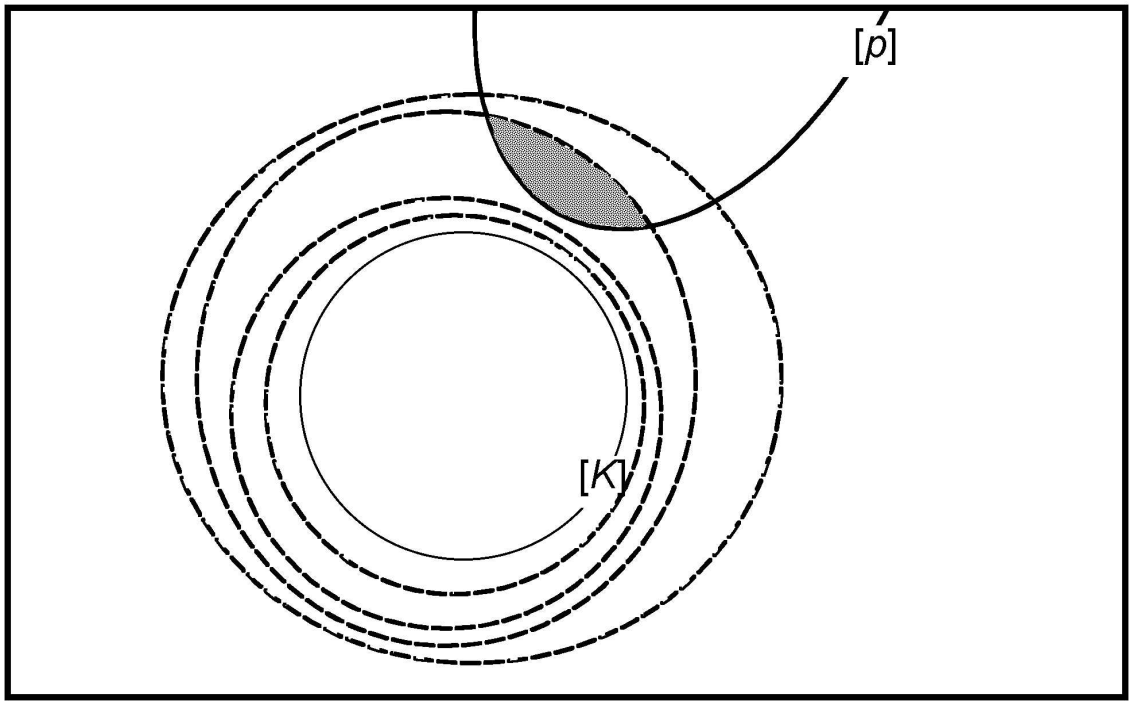
Revision—The Simplest Case

Suppose you have a set of beliefs K. Consider the possible worlds in which all these beliefs are true ($[K]$ in the picture—this is not the modal box). Accepting a new belief p that is consistent with K, amounts to taking the intersection $[K] \cap [p]$ (where $[p]$ is the set of worlds where p is true).



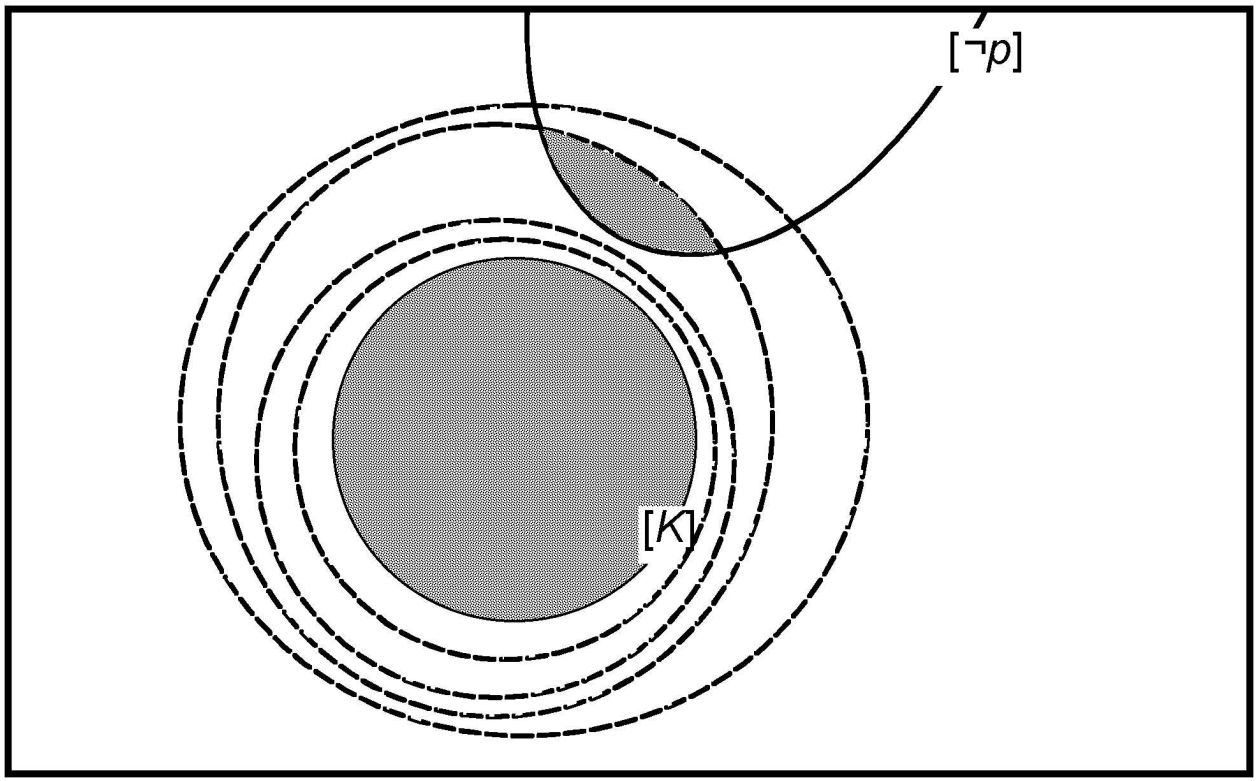
Revision—Updating with Inconsistent Information

Now suppose that p is not consistent with your existing beliefs K. What to do? The idea is to think of $[K]$ as the inner ring in a system of spheres—those beliefs that are held most strongly. Updating with p now amounts to taking the intersection of $[p]$ with the narrowest sphere around $[K]$ containing p-worlds. (This assumes that there always is such a narrowest sphere).



Contraction—Giving up a Belief

The following picture illustrates what is to be done if K entails a belief p, but p is to be given up. This boils down to taking the union of $[K]$ with the result of revising with $\neg p$.



Defaults

- Many statements are best understood as defaults. The sentence Birds fly should not be understood as 'All birds fly', but as 'Birds normally fly', or 'Birds typically fly'.
- Note that Cats have tails is consistent with a situation in which all cats have mysteriously lost their tails due to a Cat Tail Disease.
- Delgrande 1988: "The default statement $\alpha \Rightarrow \beta$ is true just when β is true in the least exceptional worlds in which α is true."
- Lewis's idea again, but now using a relation of comparative exceptionality between worlds.

Looking Back and Forward

- In this part we have seen that logics for conditional modalities can also be viewed as fragments of predicate logic.
- There are more useful logics that can be approached in this way (I'm working on the logic of common knowledge).
- The set-up automatically provides us with a single system in which many modal operators are present and it will be possible to write a theorem prover for many interacting systems.
- There are at least two pressing things on our to do list if we want the resulting system to become applicable: 1) we should combine our temporal and modal logics and 2) the logic should be extended with quantifiers. We will do this in Part 5.