PHL 232 Knowledge and Reality*

Logic Primer

This primer is a tailored introduction to the logic needed for PHL232. We start from scratch, since students are not required to take logic prior to PHL232. While some topics will be familiar to those who have taken PHL245, many others will be new to everyone. The primer provides the formal background necessary to fruitfully read and evaluate the course readings. We won't have time to cover much of this material in depth during class, so students will be expected to learn the material on their own time.

1 Arguments and their Parts

Much of what we'll do in this course involves considering and evaluating *arguments* for various positions in epistemology and metaphysics. You can think of an argument as a tool of rational persuasion: it is what you consider when you are deciding what to believe about a given topic. Here is a tidy definition:

> An *argument* is a set of claims, one of which is supported by the others

- As this definition suggests, arguments have *parts* or *components*.
  - The claim or position that an argument is designed to support is called the argument's *conclusion*.
  - The remaining parts of an argument comprise the reasons for believing the argument's conclusion. These are the argument's *premises*.

1.1 Representing Arguments: Standard Form

One of the things we'll be doing in this section is defining some of the properties an argument must have if it is to count as a *good* argument. But in order to do this, it can be helpful to have a method for representing an argument that makes its premises and conclusion, and the relation between them, explicit. In this course, we will often do this by representing arguments in *standard form*. Consider the following bit of (poor!) reasoning:

> Either there are still large caches of weapons of mass destruction lying undiscovered in Iraq despite what had been an enormous presence of international troops and officials, or the Bush administration was lying about the existence of such weapons. Either way, it is clear that no one should ever vote Republican again.

- Representing an argument in standard form involves (a) identifying the argument's premises and conclusion, and (b) representing the structure of an argument in a "list-like" way. Here is one way to represent the above argument in standard form:

P1. Either there are still large caches or weapons of mass destruction lying undiscovered in Iraq despite what had been an enormous presence of international troops and officials, or the Bush administration was lying about the existence of such weapons.

P2. If so, then no one should ever vote Republican again.

C. Therefore, no one should ever vote Republican again.

2 Properties of Deductive Arguments

An argument is a good argument just in case its premisses *support* its conclusion. What counts as support depends upon our goals. Sometimes we need the truth of an argument's premisses to somehow *guarantee* the truth of its conclusion. In other contexts we merely need the truth of the premisses to increase the *likelihood* of the conclusion's truth. To hold an argument to the first standard is to treat it as a *deductive* argument, while to hold an argument to the second is to treat it is an *inductive* argument.

- *Example*: If Obama didn't win the election, Romney did; but Romney didn't win; so Obama won the election.
- *Example*: the sun has risen every morning for the past 1000 years, so it'll rise again tomorrow. This argument is not deductive, since the sun could fail to rise tomorrow, but its premise does confer a high degree of likelihood on the argument's conclusion.

The first argument is most naturally treated as deductive, the second as inductive. Our focus in this section is going to be on the properties of (good) deductive arguments.

2.1 Validity

Validity is a property that a deductive argument has when its premises support its conclusion. We can define validity in a couple of different (but equivalent) ways.

> **Definition 1**. An argument is valid *if and only if* there is no possible way that the premises of the argument could be true and the argument's conclusion false.
>
> **Definition 2**. An argument is valid *if and only if* necessarily, if the argument's premises are true, so is its conclusion.

- It is important to note the the validity (or invalidity) of a given deductive argument is a matter that largely floats free of the truth or falsity of its premises and conclusion. In fact, valid arguments can have any of the following features:
    - False premises and a true conclusion
    - False premises and a false conclusion
    - True premises and a true conclusion

A very good test of your understanding of deductive validity is whether you can come up with examples of valid arguments that have one of these features. Given how we define valdity, to show that an argument is invalid one must show that it's premises *could* be true while it's conclusion is false.

2.2 Soundness

In addition to being valid, an argument can also be *sound.* We will say that

Definition A deductive argument is sound *if and only if* (a) it is valid, and (b) all of its premises are true.

Notice that soundness requires validity (but not vice versa). So an argument will be unsound if it has one or more false premise, or if it is invalid (or, trivially, if both conditions are met).

2.2.1 Soundness and Validity: Exercises

For each of the examples that follow, determine whether the argument is sound and/or valid.

Example 1

1. Some Jane Austen fans read George Eliot
2. George Eliot wrote *Middlemarch*
3. *Conclusion*: Some Jane Austen fans read *Middlemarch*

Example 2

1. It snowed on Labour Day
2. It didn't snow the first Monday in September
3. *Conclusion*: Toronto is larger than Hamilton

Example 3

1. Toronto is west of Kingston
2. Kingston is west of Montreal
3. Montreal is west of Ottawa
4. *Conclusion*: Toronto is west of Ottawa

Example 4

1. Cauchy hated mathematics
2. Frege wrote the *Begriffsschrift* in 1979
3. *Conclusion*: $2 + 2 = 4$

3 Propositional Logic

Many of the concepts and arguments that we will be looking at in this course are very complex, and you will find that thinking about them semi-formally can be a useful aid to understanding. In this section, we will learn about a very basic formal language—the language of *propositional logic* (PL)—that will help us do this. Some of you may already be familiar with propositional logic from other courses. If not, everything you will need to understand for our purposes is covered here.

3.1 The Language PL

3.1.1 Syntax

The basic vocabulary of the language of PL has three main components.

1. *Propositional Variables.* These are symbols used to represent or stand for simple sentences of natural language. They are variable symbols in the sense that any propositional variable may be assigned to represent any simple (declarative) natural language sentence (e.g. sentences in English).

   We will use upper-case letters of the English alphabet, such as 'P,' 'Q,' 'R,' 'S,' 'T,' etc., as our propositional variables. *Example*: we might use 'P' to stand for the sentence 'Penguins dance disco at dawn' or 'The present King of France is bald'.

2. *Logical Connectives.* The logical connectives of PL represent logical relations that hold *between* different sentences of natural language.

   We will be using each of the following connectives: '$\neg$,' '$\rightarrow$,' '$\wedge$,' '$\vee$,' '$\leftrightarrow$' (more on their interpretation below)

3. *Parentheses* (,), [,], {, }, etc. Parentheses do important work keeping the things we want to be able to say in our language readable and in good order. If you understand why parentheses are important in keeping the symbolic notation of mathematics in good order, then you already understand the role of parentheses in our language PL.

An example will bring out the importance of proper use of parentheses. The sentence 'Dogs are gluttons and cats are fiends or ferrets are sneaky' is ambiguous between two readings: ((Dogs are gluttons and cats are fiends) or (ferrets are sneaky)) *vs.* ((Dogs are gluttons) and (cats are fiends or ferrets are sneaky)). These two sentences have different meanings, since the first could be true while the second false (i.e. if ferrets are sneaky, but dogs aren't gluttons).

Armed with these syntactic components, we can define what counts as a sentence (or, more properly, a *well-formed formula* or wff) within PL. To keep this short, we'll assume that PL contains just two connectives: '$\neg$' and '$\rightarrow$'.

1. All propositional variables are wffs
2. If ϕ is a wff, then $\ulcorner\neg\phi\urcorner$ is a wff
3. If ϕ and ψ are wffs, then $\ulcorner(\phi \rightarrow \psi)\urcorner$ is a wff
4. Nothing else is a wff of PL

3.1.2 *Aside*: The Use/Mention Distinction

Attentive readers will notice that in the last section we began to use odd symbols. The Greek letters are *meta-linguistic variables* that range over expressions (in this case, expressions from PL). The quotation marks are *corner-quotes* (or quasi-quotes), and should be read as follows: '$\ulcorner(\phi \rightarrow \psi)\urcorner$ is a wff' says that *the left-to-right concatenation of '(', the sentence denoted by 'ϕ', '$\rightarrow$', the sentence denoted by 'ψ', and ')' is a well-formed formula.*

Why bother with these new bits of notation? The answer has its roots in the central (but often abused) distinction between *using* an expression and merely *mentioning* it. For example, someone who says 'John, if you name the child 'John' it will scar him' deploys the name 'John' in two very different ways: in the first, she uses it to refer to a particular individual (i.e. John); in the second, she uses quotation marks to make the name stand for itself (i.e. for the expression 'John' rather than John himself). In the first case she *uses* the name, in the second she merely *mentions* it. As the example brings out, we standardly use quotation marks as a means to distinguish mention from use.

While ordinary quotation allows us to mention, and thus say things about, particular sentences in a language, it doesn't suffice for all our expressive needs. Sometimes we need to make general claims that apply to any sentence with a given form or structure. Take 2 from the definition of wffs for PL. We cannot express this claim using ordinary quotation and propositional variables: 'If P is a wff, then '$\neg$P' is a wff' only tells us about a particular sentence P, not about sentences of PL more generally. To solve this problem, we first introduce meta-linguistic variables (standardly represented by lower-case Greek letters). 'If ϕ is a wff, then '$\neg\phi$' is a wff' improves on the previous version, since it now concerns arbitrary sentences of PL (rather than some particular sentence 'P'). Yet this revised version of 2 still falls short of what we want. '$\neg\phi$', as an instance of ordinary quotation, stands for a sentence that includes a meta-linguistic variable. But PL does not include meta-linguistic variables within its wffs, so our revised version of 2 is false. Corner-quotes solve this problem, since they serve to concatenate not the meta-linguistic variables themselves, but the sentences that these variables pick out or denote. So we may read 'If ϕ is a wff, then $\ulcorner\neg\phi\urcorner$ is a wff' as *if some arbitrary sentence of PL is a wff, then the sentence formed by concatenating '$\neg$' and the original sentence is also a wff.*

Upshot: use metalinguistic variables and corner-quotes to generalise about sentences in a given language. That said, in §7 we forgo corner-quotes, and instead use expressions autonymously (i.e. as standing for themselves). Indeed, many authors fail to adhere to strict use-mention convertions, so it is important to be vigilant when reading.

3.1.3 The Basic Semantics for PL

Semantics is the study of meaning. To give a semantics for a language requires that we both specify the meanings of our primitive or basic expressions, and formulate rules that govern how the meaning of a complex expression is determined by the meanings of its constituent expressions. So in PL, we want to know, for example, how the meaning of a complex sentence of the form $\ulcorner(\phi \wedge \psi)\urcorner$ depends upon the meanings of ϕ and ψ. Since our interest lies with the construction of valid arguments, the standard semantics for PL assigns truth-values (i.e. True = T; False = F) as meanings to both simple and complex sentences. So an adequate semantics for PL must explain how the truth-value of a complex sentence depends upon the truth-values of its constituent expressions.

We can represent the meaning of each of our five logical connectives in terms of a truth-table.

1. '$\neg$'. This is the PL symbol for *negation*. It symbolizes *not*, and English cognates such as *it is false that*, *it is not true that*, *it is not the case that*, etc. Here is its characteristic truth-table:

ϕ	$\ulcorner\neg\phi\urcorner$
T	F
F	T

2. '$\wedge$'. This is the PL symbol for *conjunction*. It symbolizes *and*, and its English cognates such as *also*, *in addition to*, *as well as*, and *but*. In a conjunction of the form $\ulcorner(\phi \wedge \psi)\urcorner$, ϕ and ψ are called *conjuncts*. Here is its truth-table:

ϕ	ψ	$\ulcorner(\phi \wedge \psi)\urcorner$
T	T	T
T	F	F
F	T	F
F	F	F

3. '$\vee$'. This is the PL symbol for *disjunction*. It symbolizes *or* (or rather, the *inclusive* use of 'or': compare 'Do you want milk or sugar?', which uses 'or' inclusively, and 'Should we go left or right?', which uses 'or' exclusively). In a disjunction of the form $\ulcorner(\phi \vee \psi)\urcorner$, ϕ and ψ are called *disjuncts*. Here is the truth-table for the $\vee$:

ϕ	ψ	$\ulcorner(\phi \vee \psi)\urcorner$
T	T	T
T	F	T
F	T	T
F	F	F

4. '$\rightarrow$'. This is the PL symbol for the *material conditional*. It symbolizes *if...then...* (as well as *...only if...*). In a conditional statement of the form $\ulcorner(\phi \rightarrow \psi)\urcorner$, ϕ is called the *antecedent* and ψ the *consequent*. Here is the truth-table for $\rightarrow$:

ϕ	ψ	$\ulcorner(\phi \rightarrow \psi)\urcorner$
T	T	T
T	F	F
F	T	T
F	F	T

5. '$\leftrightarrow$.' This is the PL symbol for the *biconditional*. It symbolizes *...if and only if...* (as well as *...just if...* and *...iff...*). Its truth-table looks like this:

ϕ	ψ	$\ulcorner(\phi \leftrightarrow \psi)\urcorner$
T	T	T
T	F	F
F	T	F
F	F	T

3.2 Sidebar on Formal vs. Natural Language

At this point you are likely wondering why we insist you learn all of this formal material—validity, soundness, and all of the wacky symbols. Most of this material you need to know in order to understand the readings for this course (though sometimes the symbols used by the authors will differ from those we use here – formalisms go in and out of fashion). For instance, truth-tables and the notion of validity help us evaluate the sceptical argument, while counterfactuals (which we'll discuss in §5) are central to Nozick's theory of knowledge and Lewis's account of causation.

We also think it important that you become acquainted with rudimentary formal languages. We can introduce formal languages via a contrast with *natural* language. English is a natural language; so is French, and Urdu, and Swahili. Natural languages arise not through definition, but through our use of them in communication (e.g. when you tell your friends about the latest lecture of PHL232) or action (e.g. couples can say 'I do' in order to complete an action – marriage). Our use of natural language is messy, context-dependent, sometimes arbitrary, and often non-literal (e.g., metaphor).

Formal languages are primarily devices that philosophers, semanticists, and mathematicians (among others) use to isolate particular aspects of natural language. Mathematicians and logicians develop formal languages to capture all and only those elements of natural language relevant to the construction of adequate proofs (a project most famously pursued by Frege 1879). As we've seen, logicians developed PL because of an interest in the use of declarative sentences within inference. So PL abstracts away from features of these sentences that do not contribute to their truth or falsity (so, for example, PL is blind to the differences between how we use 'but' and 'and').

Semanticists—philosophers and linguistics who study linguistic meaning—tend to use more complex formal languages. They use formal languages as a tool to *study* natural language. You might think of a formal language as a kind of laboratory setting: semanticists use a formal language to control for a number of variables that, if they stuck to natural language on its own, would undermine attempts to generate useful results about natural language. Philosophers also treat formal languages as laboratory settings, usually because they wish to abstract from idiosyncratic features of natural language use (a strategy made famous by Russell 1905). For example, epistemologists who study certain properties of knowledge (and cognate concepts) use a formal language in order to isolate their investigations from being infected by non-standard uses of the 'knowledge' and 'knows'. We may sometimes use 'know' in a way that doesn't mandate truth, but epistemologists abstract from this sort of use when they introduce a formal language for representing claims about knowledge.

Formal languages also possess non-instrumental interest. With some ingenuity, we can start to prove facts about formal languages. Thus we might ask: could any formalisation of arithmetic, along with associated inference rules, provide the resources to prove all and only the truths of arithmetic? Surprisingly, the answer to this question (or rather, a more careful variant of it) is *no*. Kurt Gödel, a famous Austrian mathematician and logician, established the result when he proved his famous *Incompleteness Theorems*.

The upshot: logic (and formal languages more generally) can help you become a better philosopher, and will certainly expand the class of philosophy papers you could fruitfully read.

4 Properties of Relations

Objects bear relations to one another: two people can love one another; members of a family are bound by kinship relations (e.g. siblings, parents and children, cousins); and ordinary objects stand in spatial relations to each other (e.g. α is to the right of β, or β is ten feet away from α, or α and β are co-located).

- We can represent relations using some handy formal notation (this should be familiar from mathematics):

Definition: Sentences of the form $\ulcorner R(\alpha_1, ..., \alpha_n)\urcorner$ say that a relation R relates n things (i.e. α_1 to α_n).

The number of things a relation relates is called the relation's *adicity* (or *arity*). So in our definition, the adicity of R is n. For example, the relation expressed by 'x is a brother of y' can only ever relate two things (i.e. an individual and his or her brother), and so possesses an adicity of 2. Importantly, the adicity of a relation is not affected by whether the objects it relates are identical or distinct. The relation of identity (notoriously) has an adicity of 2, despite the fact that it only ever relates objects to themselves.[1] Futhermore, relations relate objects in an *order*: hence $\ulcorner R(\alpha_1, ..., \alpha_n)\urcorner$ expresses a different state of affairs than $\ulcorner R(\alpha_n, ..., \alpha_1)\urcorner$, even though these states of affairs involve the same entities.

With our formal notation in hand, we have the means to express some interesting formal properties that relations can possess. These properties are most easily expressed if we restrict ourselves to relations with an adicity of 2 (i.e. 'dyadic' relations).

Symmetry Given a dyadic relation R, R is *symmetric* iff for all x and y, $R(x, y)$ obtains only if $R(y, x)$ does.

Transitivity Given a dyadic relation R, R is *transitive* iff for all x, y, and z, if $R(x, y)$ and $R(y, z)$ obtain then $R(x, z)$ does.

Reflexivity Given a dyadic relation R, R is *reflexive* iff for all x, $R(x, x)$ obtains

A relation that is symmetric, transitive, and reflexive is an *equivalence relation.* Some examples will help:

1. The relation expressed by 'x is a sibling of y' is symmetric (if you are someone's sibling, that person is also your sibling), but neither transitive (your sibling could be a half-sibling, and she could have a half-sibling that isn't your sibling) nor reflexive (you are not your own sibling).

2. The relation expressed by 'x is taller than y' is neither symmetric (if you are taller than someone else, that person is not taller than you) nor reflexive (you cannot be taller than yourself), but *is* transitive (if Siobhan is taller than Subha, and Subha is taller than Samia, then Siobhan is also taller than Samia).

3. Identity is an equivalence relation.

As an exercise, try to think of some familiar relations, and work out whether they are transitive, symmetric, or reflexive. It often requires some imagination to think of suitable counterexamples.

[1] Not everyone agrees that identity has an adicity of 2. Wittgenstein (1921) denied that identity deserves to be called a relation, in part because it never manages to relate two distinct objects. For a recent defence of this view, see Wehmeier (2012).

These properties of relations play an important role in a number of the topics we address in the course. For example, many philosophers think that causation must be transitive (so if A causes B, and B causes C, then A causes C). So a right account of causation must respect its transitivity (we'll see the importance of this point when we study Lewis on causation). To take another example, we'll see that distinct systems of modal logic (i.e. the logic that governs talk of necessity and possibility) emerge once we ask whether the so-called *accessibility relation* is transitive, symmetric, or reflexive.

4.1 Exercises for Relations

As an exercise, determine which of the following relations are transitive, symmetric, or reflexive (or any combination of these three).

1. x loves y
2. x eats y
3. x is a child of y
4. x is earlier than y
5. x went to school with y
6. x is similar to y
7. x overlaps (i.e. shares a part with) y
8. x taught y

5 Counterfactuals

Counterfactuals are conditional statements about what would be the case if something else were the case. For example, it seems plausible to think that had Dominic and Adam both been sick on our most recent day of class, the lecture *would have been* cancelled. This is a counterfactual claim: it is a claim about what would have been the case if Dom and Adam had been sick.

- Symbolically, we represent counterfactual conditionals as follows:

  $$\phi \mathrel{\Box\!\!\rightarrow} \psi$$

  (In English: *if it had been the case that* ϕ, *it would have been the case that* ψ)

- The logic of counterfactuals plays an important role in two of the topics we will be discussing in this course: truth-tracking theories of knowledge in epistemology, and counterfactual theories of causation in metaphysics.

- In this section, we will look at some of the logical properties of counterfactuals, and a well-known theory of their semantics due to Stalnaker (1968) and Lewis (1973b).

5.1 The Logic of Counterfactuals

The meaning of an indicative conditional statement (which, for our purposes, we'll treat as equivalent to a material conditional statement of the form $\ulcorner\phi \rightarrow \psi\urcorner$) is very different from the meaning of the corresponding counterfactual $\phi \boxrightarrow \psi$. For example, an indicative conditional is true just in case either the antecedent is false or the consequent is true (consider the truth-table for $\rightarrow$ to verify this). By contrast, a counterfactual of the form $\ulcorner\phi \boxrightarrow \psi\urcorner$ may be false when ψ is true, or when both ϕ and ψ are false.

- To begin to see why, consider the difference between 1_A and 1_B :[2]

  1_A. If Oswald didn't kill Kennedy, then someone did.
  $(\neg O \rightarrow K)$ [✓]

  1_B. If Oswald hadn't killed Kennedy, then someone would have.
  $(\neg O \boxrightarrow K)$ [x]

- 1_A is true, since Kennedy is in fact dead (so if it wasn't Oswald that did it, it was clearly *someone else* that killed him).

- But 1_B is intuitively false. Assuming Oswald acted alone, then if Oswald had stayed put in the Soviet Union, presumably no one else would have killed Kennedy.

For similar reasons

2_A. If Dom and Adam aren't teaching PHL 232 this summer, then Oswald is teaching the course

comes out true, given the truth-table for the $\rightarrow$ (ensure that you understand why). But the corresponding counterfactual

2_B. If Dom and Adam hadn't been the instructors for PHl 232 this summer, then Oswald would have been

is obviously false. So indicative and counterfactual conditionals differ considerably in what they mean. They also differ logically, as we shall see in the next section.

5.2 Logical Differences

5.2.1 Strengthening the Antecedent

- Indicative conditionals allow for *antecedent strengthening* but counterfactuals do not. Consider:

  3_A. If it is pleasant outside then Adam is teaching the class on counterfactuals this term $(P \rightarrow A)$

  3_B. $\therefore$ If it is pleasant outside and Rob Ford is misbehaving as usual, then Adam is teaching the class on counterfactuals this term $((P \wedge R) \rightarrow A)$ [✓]

  4_A. If Oswald hadn't killed Kennedy, Kennedy would have pulled the U.S. out of Vietnam $(\neg O \boxrightarrow F)$

  4_B. Therefore: If Oswald hadn't killed Kennedy and there had been a second shooter on the grassy knoll, Kennedy would have pulled the U.S. out of Vietnam $(\neg O \wedge G \boxrightarrow F)$ [x]

5.2.2 Contraposition

- Contraposition $((\phi \rightarrow \psi) \iff (\neg\psi \rightarrow \neg\phi))$ also fails for counterfactuals, though it is valid for indicative conditionals (ensure that you can use the truth-table for $\rightarrow$ to explain why!). Consider:

  5_A. If Rob Ford is misbehaving, he'll be on the news tonight.
  $(R \rightarrow N)$

  5_B. $\therefore$ If Rob Ford isn't on the news tonight, he hasn't been misbehaving.
  $(\neg N \rightarrow \neg R)$ [✓]

  6_A. If Oswald hadn't killed Kennedy, Johnson wouldn't have been President.
  $(\neg O \boxrightarrow \neg J)$

  6_B. Therefore: If Johnson had been President, Oswald would have killed Kennedy.
  $(J \boxrightarrow O)$ [x]

5.2.3 Transitivity

- Hypothetical syllogism $(((\phi \rightarrow \psi) \wedge (\psi \rightarrow \chi)) \rightarrow (\phi \rightarrow \chi))$ is valid for indicative conditionals. This is because material implication is transitive. But counterfactual implication is *intransitive*. Consider $7_A - 7_C$:

  7_A. If Adam goes to the party Dom will stay home.

  7_B. If Dom stays home, no-one will have a good time.

  7_C. $\therefore$ If Adam goes to the party no-one will have a good time.

- This is clearly a valid argument. By contrast, consider

  8_A. If I had stayed home the party would have gone all night (because parties are livelier when I'm not around).

[2] These classic examples were introduced by Adams (1970).

8_B. If the party had gone all night you would have had a headache in the morning (you like to stay at parties until the bitter end).

8_C. $\therefore$ If I had stayed home you would have had a headache in the morning.

- This argument is invalid (perhaps you avoid parties that I don't attend).

5.3 A Similarity-Based Semantics for Counterfactuals

Stalnaker (1968) and Lewis (1973b) developed a possible worlds-based semantics for counterfactual conditionals that explains the logical differences between indicative conditionals and counterfactuals that we have noted. We will be looking in more detail later at the concept of a possible world. But for present purposes, it is enough to think of a possible world as an alternative way things could have been.

- The basic idea behind the Stalnaker-Lewis approach is to distinguish between *degrees of divergence* from actuality: one possible world w may be *more similar*, overall, to the way things actually are than some other possible world w^*. In such a case, we say that w is "closer" to actuality than w^* is. For example, a world that is just like the way things actually are, except for the fact that I am sitting down now, is closer to actuality than a world at which I am both sitting down now and the only person that exists.

- Here is the core thesis of the similarity-based semantics for counterfactuals proposed by Lewis and Stalnaker:

  $\ulcorner \phi \mathrel{\square\!\!\rightarrow} \psi \urcorner$ is true at a world w *if and only if* :

  (i) ψ is true at some ϕ-world w' (where a ϕ-world is just one in which ϕ is true); and

  (ii) w' is closer to actuality than any ϕ and $\neg\psi$ world w''

- This account *explains why antecedent strengthening fails*. A world at which Oswald doesn't shoot but there is a second shooter on the grassy knoll is less similar, overall, to actuality than a world at which Oswald simply doesn't shoot.[3]

- It also *explains why contraposition fails*. The closest worlds where Oswald doesn't shoot are worlds where Kennedy finishes out his term in office (and hence worlds in which Johnson is not President in '63). But the closest worlds where Johnson is President in '63 are worlds where Kennedy is doing something different with his life (and not incurring the ire of Oswald).

- The account also *explains why transitivity fails*. The closest worlds where I stay home are worlds where the party goes all night. And the closest worlds where the party goes all night are worlds where you've got a headache in the morning. But the closest worlds where I stay home are worlds where you don't want to party (and so you avoid a morning headache).

[3] Assuming the second shooter conspiracy theory is actually false.

5.4 More on Indicative vs. Counterfactual Conditionals

You've now been taught the basic difference between indicative and counterfactual conditionals.[4] To test your grasp of this distinction, we've put together a number of exercises. Some of these require you to determine the soundness and/or validity of arguments, others just require that you determine the truth value of a sentence that contains one or more conditionals. We start with the sentences:

1. Had Harry Potter not been the child of prophecy, Neville Longbottom would have been.
2. If it rains, it is cloudy
3. If it were to be cloudy, it would rain
4. If Shakespeare didn't write Hamlet, someone else did
5. If Homer Simpson is a cartoon character, he is an animated character
6. If you fail the exam, you fail the class
7. If you fail the class, you fail the exam
8. If you pass the exam, the long paper, and one of the short papers, you pass the class
9. Had Hilary Clinton won the Democratic Primary in 2008, either she or McCain would have won the election

Example 1

1. Were at least some arithmetical statements true (e.g. '3 + 7 = 10'), numbers would exist
2. At least some arithmetical statements are true
3. *Conclusion*: numbers exist

Note: this is akin to the main argument in Frege (1884); Field (1980) accepts 1 but denies 2.

Example 2

[4] The literature on conditionals is enormous, and gets larger every year. Philosophers and linguists since Stalnaker (1968) and Lewis (1973b) have extended or revised their basic observations in surprisingly and sophisticated ways (for example, Edgington 1995 argues that conditionals do not have truth conditions, and so lack truth values). For a survey of the major literature, see Bennett (2003).

1. I can conceive of a scenario in which my mind exists without my brain
2. If my mind were to exist without my brain, my mind would not be identical to my brain
3. *Conclusion*: my mind is not identical to my brain

6 Closure Principles and Epistemic Logic

Our formulation of the traditional sceptical argument exploits an (apparent) fact about knowledge, namely that it is closed under known entailment. What does this mean? The relevant premiss says that if S knows that if she has hands then she is not a brain in a vat, then if S knows that she has hands, S knows that she is not a brain in a vat. This premiss is an instance of the more general claim that knowledge is closed under known entailment:

KE-Closure If S knows that if p then q, then if S knows that p, S knows that q.

This principle a bit of a mouthful. But it enjoys some strong intuitive support. Imagine what would happen if it weren't true: we could know that q follows from p, and know that p, but still fail to know that q. (Nevertheless, we've seen that Nozick (1981) thinks the instance of the principle that shows up in the sceptical argument is false.)

The principle owes its name to a notion of 'closure' borrowed from logic and mathematics. A given set is closed under some operation iff the result of applying the operation to a member of the set will also be member of the set. For example, the set of natural numbers {0, 1, 2, ...} is closed under the operation of addition, since the sum of any two natural numbers will also be a natural number. Our interest lies with the set of *known* propositions. KE-Closure says that if a known proposition p is related to another proposition q by a known entailment (i.e. S knows that if p then q), q also falls within the set of known propositions. Compare K-Closure with some related closure principles:[5]

E-Closure If p then q, then if S knows that p, S knows that q

KP-Closure If S knows that q is probable given p, then if S knows that p, S knows that q

KB-Closure If S believes that if p then q, then if S knows that p, S knows that q

E-Closure – the principle that knowledge is closed under entailment – seems false, since its truth would require that we know all the consequences of everything we know. What about KP-Closure and KB-Closure? (*Hint*: remember that knowledge is factive)

[5]In addition to the closure principles cited below, Gettier (1963) also relies upon yet another closure principle.

Epistemologists have devised useful formal notations that allow them to express closure principles (and other facts about epistemic concepts such as knowledge and belief) without the messiness of natural language. Let's look at one such notation:

1. $K_S p =_{def}$ *S knows that p*
2. $B_S p =_{def}$ *S believes that p*

Given this notation, we can re-write KE-Closure and some basic facts about knowledge and belief:

$$KE - Closure: \ K_S(p \rightarrow q) \rightarrow (K_S p \rightarrow K_S q)$$

$$Factivity: \ K_S p \rightarrow p$$

$$Knowledge\ Entails\ Belief: \ K_S p \rightarrow B_S p$$

$$Knowledge\ Outstrips\ Belief: \ B_S p \not\rightarrow K_S p$$

$$Belief\ is\ Fallible: \ B_S p \not\rightarrow p$$

If you read the recommended Smithies (2012), you'll come across his discussion of Moore's paradox (due, as the name suggests, to G.E. Moore). Moore's paradox concerns an asymmetry between our self-avowals of beliefs and our attributions of belief to others. Moore observed that someone (say S) who asserts that $p \& \neg B_S p$ is somehow incoherent, whereas if S were to instead assert that $p \ \& \ \neg B_{S'} p$ (of some other person, S') she would be perfectly coherent.

One puzzle, for those who study Moore's paradox, lies in identifying what kind of incoherence Moore's observation tracks. If we assume that belief is fallible, the incoherence cannot be due to inconsistency (so Moore's paradox is not, strictly speaking, a paradox). But if we aren't inconsistent when we assert Moore-paradoxical sentences, in what sense will the assertion prove rationally incoherent?

7 Modal Logic

Our topic in this section is modal logic, or the logic of possibility and necessity. A basic understanding of modal logic will aid your understanding of many of the views we'll encounter in this course. [*Note*: to avoid clunkiness, in this section we put aside corner quotes in favour of using expressions as symbols for themselves (the so-called *antynomous* use of language).]

What we're going to do here is consider a very basic system of propositional modal logic. The vocabulary of propositional modal logic is very simple, and comprises (a) the vocabulary of PL with which you are already familiar (propositional variables, logical connectives, parentheses, etc.) and (b) two new operators: the "box" $\Box$ and the "diamond" $\Diamond$.

- The intended interpretation of $\Diamond$ is that it symbolize English language locutions like 'possibly', 'might', and so on. Whenever ϕ is a well-formed formula of propositional logic, $\Diamond\phi$ is a well-formed formula of modal logic.
- The intended interpretation of $\Box$ is that it symbolize English language locutions like 'necessarily', 'it must be the case that', and so on. Whenever ϕ is a well-formed formula of propositional logic, $\Box\phi$ is a well-formed formula of modal logic.
- Note that $\Diamond$ and $\Box$ are *duals*, and so are interdefinable in the following sense:
  - $\Diamond\phi \iff \neg\Box\neg\phi$
  - $\Box\phi \iff \neg\Diamond\neg\phi$
- These interdefinitions should make intuitive sense: what is possibly the case is just what is not necessarily false, and what is necessarily the case is what could not possibly have been otherwise.

7.1 A Basic Modal Semantics

Just as we gave a precise account of the meanings of each of our logical connectives in terms of its characteristic truth-table, it would be nice to have a precise account of the meaning of $\Diamond$ and $\Box$. But unlike $\neg, \wedge, \vee, \rightarrow$ and $\leftrightarrow$, the meanings of $\Diamond$ and $\Box$ cannot be given in terms of a truth-table. This is because unlike the logical connectives, $\Diamond$ and $\Box$ are not *truth-functional*, in the sense that the truth-value of $\Diamond\phi$ and $\Box\phi$ is not a function of the truth-value of ϕ. What we will do instead is explain the meaning of $\Diamond$ and $\Box$ in terms of a *model* for the language of propositional modal logic.[6]

- Think of a model for the language of propositional modal logic as a structure of the form $\langle \mathsf{W}, \mathsf{V}\rangle$, in which
  - W is set of possible worlds
  - V is a two-place function—the model's "valuation" function—calculating the truth-value of every formula of the language relative to each possible world in W. We will let $\mathsf{V}(\phi, w) = T$ mean that ϕ is true at or according to the world w.
- Although this might seem a little complicated, when it comes to non-modal formulas of our language the basic idea is actually nothing new: the valuation function V simply lets us represent, in a slightly different way, what you already know about the connectives in virtue of understanding their characteristic truth-tables. For example, where ϕ and ψ are any well-formed formulas (and suppressing the corner-quotes):
  - $\mathsf{V}(\neg\phi, w) = T$ *iff* $\mathsf{V}(\phi, w) = F$
  - $\mathsf{V}(\phi \wedge \psi, w) = T$ *iff* $\mathsf{V}(\phi, w) = T$ and $\mathsf{V}(\psi, w) = T$
  - $\mathsf{V}(\phi \vee \psi, w) = T$ *iff* either $\mathsf{V}(\phi, w) = T$ or $\mathsf{V}(\psi, w) = T$
  - $\mathsf{V}(\phi \rightarrow \psi, w) = T$ *iff* either $\mathsf{V}(\phi, w) = F$ or $\mathsf{V}(\psi, w) = T$
  - $\mathsf{V}(\phi \leftrightarrow \psi, w) = T$ *iff* $\mathsf{V}(\phi, w) = \mathsf{V}(\psi, w)$
- What about modal formulas of the language—formulas of the form $\Diamond\phi$ and $\Box\phi$? Here is a natural thought: intuitively, $\Diamond\phi$ should come out true just in case ϕ is true at *some* possible world, and $\Box\phi$ should come out true just in case ϕ is true at *every* possible world.
- We can make this idea more precise in terms of our basic semantics as follows:
  - $\mathsf{V}(\Diamond\phi, w) = T$ *iff for some world* w' *in* W: $\mathsf{V}(\phi, w') = T$
  - $\mathsf{V}(\Box\phi, w) = T$ *iff for all worlds* w' *in* W : $\mathsf{V}(\phi, w') = T$

7.2 Applications of Modal Logic

Modal logic can seem quite baroque. Why should we care about what it takes for sentences involving necessity and possibility to be true? Most of the time we don't really care about other ways the world could be: we want to know how things stand here in the actual world. Given the troubles that follow when we try to interpret modal discourse, we might be inclined to jettison all such talk. W. V. O. Quine adopted a version of this stance.

Unhappily for Quine, it turns out that the apparatus of modal logic has wide philosophical and technical application.[7] We've already seen, in Lewis (1973a) and Nozick (1981), how a proper treatment of modal logic brings with it a proper treatment of counterfactuals. But philosophers have also found use for the notion of a possible world when analysing belief, knowledge, justification, evidence, laws of nature, conceivability, meaning, perception, and the existence of God (to pick just a few prominent examples). These different uses tend to correspond to different ways we might restrict the set of worlds over which the modal operators are defined.

Consider the different uses of the modal expression 'might'. One use tracks what we might call *logical* possibility. Something *might* be the case in this sense if it is consistent with the laws of logic (i.e. it doesn't entail a contradiction when combined with the laws of logic). Another use, arguably more restrictive, tracks nomological possibility, where something is nomologically possible iff it could occur in a world that shares our laws of nature.

But we also use 'might' to track what could be the case given what we *know* (so-called *epistemic* possibility). Often philosophers think of inquiry – the pursuit of knowledge – as an attempt to fill-in a picture of the actual world (cf. Stalnaker

[6] This kind of semantics for modal logic was developed in the 1950's and 1960's by Kripke (1959, 1963). But in many ways Kripke's approach was anticipated by Tarski and McKinsey (1944), Carnap (1947), and others.

[7] Ironically, it was Quine's own work (in his famous Quine 1951) that opened to door to the ambitious use of modal logic in Kripke (1980) and Lewis (1986) (to pick two prominant examples).

1987). This picture becomes more complete with every new proposition we come to know. But until the picture is complete, some ignorance will persist about the exact character of the actual world (we say: 'It might be that the keys are on the desk, but I don't know that to be the case'). Modal logic can help us model this ignorance. Filling in our picture of the world becomes a matter of gradually excluding worlds from the set of those consistent with what we know (i.e. those possible worlds that would make all of our knowledge-constituting beliefs come out as true). More formally: S knows that p iff p is true in every world consistent with what S knows; p might be the case for S iff p is true in some world consistent with what S knows; if S learns that p, the set of worlds consistent with what S *now* knows is identical to the set that results when we remove the set of worlds in which p is false from the set of worlds consistent with what S *used* to know.

8 Answers

Section 2.2.1:

> 1. Invalid & Unsound (though all the premisses are true); 2. Valid, unsound; 3. Valid, unsound; 4. Valid, unsound

Section 4.1

> 1. None; 2. None; 3. None; 4. Transitive; 5. Symmetric, Reflexive; 6. Reflexive, Symmetric; 7. Reflexive, Symmetric; 8. None

Section 5.4

> 1. True; 2. True; 3. False; 4. True; 5. True; 6. False; 7. False; 8. False; 9. True; Ex1. Valid, but perhaps not sound; Ex. 2 invalid, unsound.

References

Adams, E. (1970). Subjunctive and indicative conditionals. *Foundations of Language*, 6:89–94.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press.

Carnap, R. (1947). *Meaning and Necessity*. University of Chicago Press.

Edgington, D. (1995). On conditionals. *Mind*, 104:235–329.

Field, H. (1980). *Science Without Numbers: A Defense of Nominalism*. Princeton University Press.

Frege, G. (1879). *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle.

Frege, G. (1884). *The Foundations of Arithmetic*. Northwestern University Press.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 23:121–123.

Kripke, S. (1959). A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24(4):1–14.

Kripke, S. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16:83–94.

Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.

Lewis, D. (1973a). Causation. *Journal of Philosophy*, 70:556–567.

Lewis, D. (1973b). *Counterfactuals*. Blackwell.

Lewis, D. (1986). *On The Plurality of Worlds*. Blackwell.

Nozick, R. (1981). *Philosophical Explanations*. Harvard University Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review*, 60:20–43.

Russell, B. (1905). On denoting. *Mind*, 14(56):479–493.

Smithies, D. (2012). Moore's paradox and the accessibility of justification. *Philosophy and Phenomenological Research*, 85(2):273–300.

Stalnaker, R. (1968). A theory of conditionals. In Rescher, N., editor, *Studies in Logical Theory*, volume 2 of *American Philosophical Quarterly Monograph*, pages 98–112. Blackwell.

Stalnaker, R. (1987). *Inquiry*. MIT Press.

Tarski, A. and McKinsey, J. C. C. (1944). The algebra of topology. *Annals of Mathematics*, 45:141–191.

Wehmeier, K. (2012). How to live without identity - and why. *Australasian Journal of Philosophy*, 90(4):761–777.

Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus*. Routledge.