Greg Restall, *Arché, University of St Andrews* · 2025 Wendy Huang Lectures, National Taiwan University

I have three aims for this set of lectures.

1. To give an account of what is so distinctive about *logic*, insofar as logical notions have a grip on whatever can be *said* or *thought*.
2. To clarify the connections between logic and *semantics*, the theory of meaning.
3. I will provide a philosophical motivation and non-technical introduction to the more technical work I have done in my forthcoming manuscript *Proof, Rules, and Meaning*.

https://consequently.org/w/prm

Today's topic is the fundamentals of *proof, rules,* and *meaning*. I aim to give an opinionated introduction to INFERENTIALIST SEMANTICS. ¶ In the second lecture, I will show how this inferentialist semantics applies to issues of predication, quantification, and modality, and thereby provides some distinctive insight into 'possible worlds' models for quantified modal logics. ¶ Then in the final lecture, I will relate inferentialist semantics to truth-conditional accounts of meaning for natural languages.

1.1 LOGIC & SEMANTICS / PROOFS & MODELS

What has *logic* to do with semantics? There are many possible answers:

- Logics give an account of the meanings of some concepts: the logical constants.
- Logical concepts (equivalence, entailment, contradiction, etc.) are useful when accounting for the significance of what is said.
- Logical tools provide frameworks for semantic theories. (In the 20th and 21st Centuries, truth-conditional semantic theories are chief among these.)

In these lectures, I am *primarily* exploring the third answer.

» «

Kurt Gödel's *completeness theorem* is, by one measure, the greatest result in 20th Century Logic. The completeness theorem, connecting proofs and models, shows that the concept of logical validity has *two sides*.

Kurt Gödel (1906–1978)

VALIDITY (PROOFS): $A \vdash B$ says that *there is a proof* from A to B.

VALIDITY (MODELS): $A \vDash B$ says that *there is no counterexample* to the argument from A to B. (That is, no *model* taking the premise to be *true* and the conclusion *false*.)

SOUNDNESS: If $A \vdash B$ then $A \vDash B$. (If *there is a proof* then *there is no counterexample*.)

COMPLETENESS: If $A \vDash B$ then $A \vdash B$. (If *there is no counterexample* then *there is a proof*.)

A soundness theorem states there is no *overlap* between proofs and counterexamples, and a completeness theorem states that there is no *gap*.

» «

The soundness and completeness theorems show that the concept of validity can be understood both in terms of models and in terms of proofs. It would be surprising if models were significant for semantics and proofs were not.

However, theorists have, by and large, taken model theory—formulated as it is, in terms of mappings from linguistic items to an extra-linguistic structure, so easily interpreted as a *reference* relation between language and world—as properly *semantics* while relegating proof theory to mere *syntax*.

For a recent example, see Wolfgang Schwarz's recent review of Andrew Bacon's *Philosophical Introduction to Higher-order Logics*. https://www.umsu.de/blog/2025/821.

My aim in this series of lectures is to *explain* and to *exploit* the semantic significance of proof. I will provide a guide to a rigorous *inferentialist* semantics, and then I will *use* that semantics to give a fresh perspective on possible worlds models and truth-conditional semantics for natural language.

1.2 QUESTIONS & ANSWERS / NATURAL DEDUCTION

Let's begin with the first topic—giving an account of what is distinctive about *logic*, insofar as logical notions apply to whatever can be said or thought. What *can* be thought or said? ¶ We start, as usual, by considering *declarative* thoughts and sentences: the expressions that we use to make *claims*.

Though, with Nuel Belnap, we should remember that *declaratives are not enough* [2].

There are many ways to study declaratives. We won't start by looking at the *structure* declaratives might have. First, there are some generic features of the class of declarative judgements *as such* which can be isolated, before we turn to their structure.

Nuel Belnap (1930–2024)

When I make a declarative claim or think a declarative thought *I take a stand on an issue*. ¶ This is an issue about which others, at least potentially, might *disagree*.

If I take it that Queensland will win the Sheffield Shield next year, you might very well disagree. Or, you might not be so confident as me and withhold judgement, or you might not be in a position to even formulate the thought or understand the claim.

See `https://en.wikipedia.org/wiki/Sheffield_Shield`

The items we take to be the constituents of our proofs as premises and conclusions are the kinds of things we can claim, and question, and disagree about, whatever their structure. ¶ How are we to characterise this *disagreement*?

One way is to look at a part of speech paired with the declarative: the polar question. For any claim that A, there is the polar question $A^?$, asking whether A.

To make the claim that A, I take it, has the same upshot as answering *yes* to the question $A^?$. We'll call that speech act *asserting* A. ¶ To say *no* to $A^?$ is to *deny* A.

For more on polar questions, their answers, and how they relate to assertion and denial see my paper 'Questions, Justification Requests, Inference, and Definition' [10].

The answers *yes* and *no* to $A^?$ *disagree*. They take opposing sides on the issue. ¶ The answer *no* stands opposed to *yes*, and conversely, *yes* stands opposed to *no*. ¶ This opposition is not *causal*—my saying *no* does not *prevent* your saying *yes*, or even prevent *my* saying *yes*—it is, in some sense, *normative*.

Given that the aim of this exercise is to settle issues, if I have ruled out one answer to the polar question $A^?$ the answer that remains is the other.

» «

None of this is to say anything about *proof*, and properly logical notions, such as entailment, contradiction, and equivalence.

A proof is a specific species of the genus *answers to justification requests*. These are nearly as fundamental to the notion of making a *claim* as are polar questions. Whenever an assertion is made, a possible response for a hearer is to demur and ask for justification.

If I say that Queensland will win the Sheffield Shield, you may respond: 'Really!? Why?' This indicates that you don't accept my claim, and you ask me to support it.

To *prove* A, given some background context (in which certain given claims are taken for granted) is to *show* that A, in a certain rigorous gap-free sense. Some things follow from this thin characterisation:

'Gap-free' in *what* sense, exactly? Good question. Hold that thought for now. This will be clarified as we go along.

First, proofs can be chained together.

If we have a proof of A from some context $\mathcal{C}$, and we can prove B from $\mathcal{C}', A$ (another context, in which we take A for granted, among other things), then we can prove B from the context $\mathcal{C}, \mathcal{C}'$ by proving B from $\mathcal{C}', A$ (thereby meeting the request to justify B, but incurring the cost of appealing to A), and then using the proof from $\mathcal{C}$ to justify the use of A, on the basis of the background assumptions in $\mathcal{C}'$.

The two proofs π and π' are chained together like this:

$$\begin{array}{c} \mathcal{C} \\ \vdots\,\pi \\ \mathcal{C}',\ A \\ \vdots\,\pi' \\ B \end{array}$$

Proofs, whatever they are, can be chained together. Whatever we take for granted in some context can be justified in another. Proofs, understood in this way, lead from the background context to the conclusion, in a step-by-step fashion.

Chaining proofs together is of little use unless we have some proofs to begin with. ¶ This is the other point at which the background context $\mathcal{C}$ can play a role. If the formula A is already taken for granted in $\mathcal{C}$, we need do nothing else to justify it *in that context*. It is simply taken as given.

We depict the proof of A in a context in which A is taken as given, like this:

$$A$$

Using the notation from before, writing $\mathcal{C} \vdash A$ to represent the existence of a proof of A from the context $\mathcal{C}$, we have the following two *structural* features of proof:

- If A is in $\mathcal{C}$ then $\mathcal{C} \vdash A$
- If $\mathcal{C} \vdash A$ and $\mathcal{C}', A \vdash B$ then $\mathcal{C}, \mathcal{C}' \vdash B$.

» «

What about the polarity between assertion and denial? ¶ A context in which I say *yes* and I say *no* to the same claim is, in some sense, defective. ¶ We will represent *showing* that a context $\mathcal{C}$ is defective by expanding our notion of proof like so.

$$\frac{A \qquad \cancel{A}}{\sharp}$$

A proof ending in '$\sharp$' shows that our context is self-undermining. More generally, if I have a proof from a starting context $\mathcal{C}$ to A, I can extend this into a proof that leads us from $\mathcal{C}, \cancel{A}$ to $\sharp$. We can use the rule to extend a proof in this way:

$$\begin{array}{c} \mathcal{C} \\ \vdots\,\pi \\ \dfrac{A \qquad \cancel{A}}{\sharp}\,\lightning \end{array}$$

A proof that ends in $\sharp$ does not 'prove $\sharp$'. It is a demonstration that the starting context is out of bounds. The fundamental idea is that a context containing A (*taking A for granted*) and $\cancel{A}$ (*ruling A out*) is out of bounds.

So, using the turnstile notation, we have $A, \cancel{A} \vdash \sharp$. The context consisting of an assertion and denial of A is defective, or out of bounds.

In a context in which A is granted, adding a *denial* of A is ruled out. Equally, in a context in which A is denied, going on to *grant* A is ruled out. ¶ On the other hand, if I *try* to deny A, and it turns out that this is ruled out, then the context implicitly settles the question $A^?$: it is undeniable. ¶ This grounds the following rule:

$$\begin{array}{c} \mathcal{C}\ [\cancel{A}]^i \\ \vdots\,\pi \\ \dfrac{\sharp}{A}\,{\downarrow}^i \end{array}$$

If the proof π reduces the context $\mathcal{C}$—with the temporary addition of the denial of A—to absurdity, then the context $\mathcal{C}$—now with that denial *discharged*—settles A positively, since ruling it out is excluded.

» «

We can think of the contexts $\mathcal{C}$ in which certain claims are ruled in and other claims are ruled out as forming the different POSITIONS we are able to formulate using the conceptual resources of the language at hand. ¶ In any language equipped with a notion of proof like this, we will say that a position $\mathcal{C}$ is OUT OF BOUNDS if there is a proof from $\mathcal{C}$ to $\sharp$. (So, $A, \cancel{A}$ is, at the very least, out of bounds.) ¶ We will say that a position is AVAILABLE if it is not out of bounds.

» «

That's the bacgkround context in which we formulate proofs. We have said nothing yet about the structure of the judgements that populate our proofs. It is to the behaviour of *connectives* that we now turn.

Here are the basic inference rules governing conjunction, from Gentzen's pioneering work on natural deduction [4], popularised and systematised by Dag Prawitz [6].

$$\frac{A \qquad B}{A \wedge B}\,{\wedge}I \qquad \frac{A \wedge B}{A}\,{\wedge}E \qquad \frac{A \wedge B}{B}\,{\wedge}E$$


Gerhard Gentzen (1909–1945)

The inference rules for conjunction govern *conjunction*, and nothing else. We are told what it takes to *prove* a conjunction (namely, both the conjuncts) and what you can prove *from* the conjunction (namely, either conjunct). ¶ These rules look as basic as one could hope for. If I am asked why it is the case that $A \wedge B$, and I'm given a reason for A and a reason for B, I don't seem to need anything *else*, above A and above B to justify $A \wedge B$. ¶ Conversely, if I am already in a position to grant the conjunction $A \wedge B$, I need look for no *further* justification to answer the question A, or B. These inferences seem immediate, direct, and as gap-free as one could want. ¶ Similar rules can be given for the conditional, and for negation.

$$\dfrac{\begin{matrix}[A]^i\\ \vdots\\ B\end{matrix}}{A \to B}{\scriptstyle \to I^i} \qquad \dfrac{A \to B \quad A}{B}{\scriptstyle \to E} \qquad \dfrac{\begin{matrix}[A]^i\\ \vdots\\ \sharp\end{matrix}}{\neg A}{\scriptstyle \neg I^i} \qquad \dfrac{\neg A \quad A}{\sharp}{\scriptstyle \neg E}$$

We can combine these rules, to formulate longer proofs. ¶ The first is a proof from $\neg\neg p$ to p. The second, from $\neg(p \to q)$ to $p \wedge \neg q$.

$$\dfrac{\dfrac{\neg\neg p \quad \dfrac{\dfrac{[p]^1 \quad [\not p]^2}{\sharp}{\scriptstyle \lightning}}{\neg p}{\scriptstyle \neg I^1}}{\sharp}{\scriptstyle \neg E}}{p}{\scriptstyle \downarrow^2} \qquad \dfrac{\dfrac{\dfrac{\neg(p \to q) \quad \dfrac{\dfrac{\dfrac{[p]^1 \quad [\not p]^2}{\sharp}{\scriptstyle \lightning}}{q}{\scriptstyle \downarrow}}{p \to q}{\scriptstyle \to I^1}}{\sharp}{\scriptstyle \neg E}}{p}{\scriptstyle \downarrow^2} \quad \dfrac{\dfrac{\neg(p \to q) \quad \dfrac{[q]^3}{p \to q}{\scriptstyle \to I}}{\sharp}{\scriptstyle \neg E}}{\neg q}{\scriptstyle \neg I^3}}{p \wedge \neg q}{\scriptstyle \wedge I}$$

In the second proof, note that the rightmost $\to I$ inference is *vacuous*: zero instances of the assumption p are discharged in the step from q to $p \to q$. Similarly, the first $\downarrow$ inference is vacuous, in that the discharged denial q occurs zero times in the context. This proof system is unapologetically *irrelevant*.

These rules suffice for classical propositional logic [9], given the bilateral setting in which the context can contain both positive judgements (A) and negative ones ($\not A$).

» «

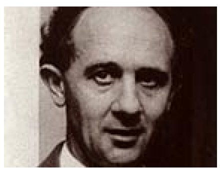
It is compelling to think of the introduction and elimination rules for logical concepts as, in some sense, *definitions* of the concepts they govern. ¶ Someone who says, "I grant A and grant B but I'm not sure about $A \wedge B$", is, in some sense, missing the point, and not grasping the concept '$\wedge$'. ¶ It would be a mistake to look for *another premise* to add to A and to B to justify the inference to $A \wedge B$ if our interlocutor does not make the inference on their own. ¶ Facility with the introduction and elimination rules seems like a criterion for understanding the concept '$\wedge$'.

Taking certain inference rules to be definitional in this sense provides one possible response to the Tortoise's behavior in Lewis Carroll's classic paper [3, 10].

Can we take this thought *seriously*? Can we treat inference rules as definitions of the concepts they govern? ¶ If we can get away with this idea, then we can answer a question I left hanging: A proof is gap-free in the sense that each of the inference steps is definitional of the concepts they govern. ¶ Similarly, the purely structural rules $\downarrow$ and $\lightning$ constrain the field of contents upon which the connective rules operate. ¶ Proofs have no gaps because conceptual competence suffices to underwrite each step.

1.3 RULES & DEFINITIONS / INFERENCE & MEANING

Taking introduction and elimination rules as *definitions* is a compelling thought, but it cannot be the end of the matter. As Arthur Prior showed in 1960 [7], if we could simply specify introduction and elimination rules for a putative concept and declare them to be *definitions*, we could prove anything we like. ¶ Prior introduced tonk with two rules:



Arthur Prior (1914–1969)

$$\dfrac{A}{A \text{ tonk } B}{\scriptstyle \text{tonk}I} \qquad \dfrac{A \text{ tonk } B}{B}{\scriptstyle \text{tonk}E}$$

These rules trivialise. From any premise A, you can prove any conclusion B.

$$\dfrac{\dfrac{A}{A \text{ tonk } B}{\scriptstyle \text{tonk}I}}{B}{\scriptstyle \text{tonk}E}$$

So we must either (1) explain the difference between definitions that are OK to adopt and those that aren't, or (2) say more about what counts as a definition (and find a way to exclude Prior's rules for tonk from the fold, while retaining the rules for connectives like $\wedge$), or (3) give up the 'definition' game entirely. ¶ I choose option (2).

» «

Here is one proposal for what a definition does for us: It shows us how extend a starting language $\mathcal{L}_1$ with a new item of vocabulary, to form an extended language, $\mathcal{L}_2$, in which we can use the new item in some determinate way.

An *abbreviative* definition is a paradigm case. If $\mathcal{L}_1$ is a language in which I can say 'x has sides all of the same length', then I can extend the language with the term 'equilateral', which is defined as meaning 'having sides all of the same length'. ¶ To adopt the definition is not only to take it that x is equilateral if and only if the sides of x have the same length, but *also*, to grant that a demonstration that x's sides have the same length conclusively answers the question of whether x is equilateral, and *vice versa*. ¶ Someone who took it as given that x's sides were all of the same length, but still took it to be an open question whether x is equilateral would thereby show that either they have not *adopted* the definition, or at least, they have not *mastered* it.

In the language of *proofs*, introducing a newly defined expression E in terms of previously introduced vocabulary D, can be seen as adopting this *invertible* rule:

$$\frac{\mathrm{D}}{\mathrm{E}}\;\mathrm{E}\mathit{Df}$$

In any context at all, from D you can infer E, and *vice versa*. ¶ Here, $\mathcal{L}_2$ is formed by taking the language $\mathcal{L}_1$ (with all its inference rules taken as given), adding the vocabulary item E, and constraining it by imposing the two-way rule E*Df*.

Abbreviative definitions have two useful features. ¶ They are CONSERVATIVE: positions in the original language ($\mathcal{L}_1$) that were available beforehand remain available afterward from the point of view of $\mathcal{L}_2$. ¶ The new vocabulary of $\mathcal{L}_2$ introduces a new rule, and scope for formulating different proofs in $\mathcal{L}_2$, even concerning $\mathcal{L}_1$-vocabulary alone. However, it does not introduce new facts about what is *provable* in the $\mathcal{L}_1$ language, since any $\mathcal{L}_2$-proof may be rewritten as an $\mathcal{L}_1$-proof.

In any $\mathcal{L}_2$-proof, replace every E by D, cut out the E*Df* steps, which are then redundant, and the result is an $\mathcal{L}_1$-proof.

Abbreviative definitions are also UNIQUELY DETERMINED: The newly introduced expression E is *fixed*, in the sense that if we happened to add the term *twice* (E_1 and E_2, say), using definitions of the same form, then the two defined terms are equivalent, in the sense that we can infer one from the other, in any context. ¶ This means that our definition is not merely a schematic specification which could be instantiated in different non-equivalent ways.

Consider the traditional axioms and rules governing the necessity modality in the modal logic S5. These can be instantiated in different ways in the one language. $\Box_1$ and $\Box_2$ could be two different necessity operators, each equally satisfying the traditional axioms and rules for S5. These do not define that modality, but at most, define what it is for an operator to be *an S5 necessity*.

CONSERVATIVITY and UNIQUENESS are useful features. A *conservative* definition is safe. (It is not like substantial claim that rules out some positions. No $\mathcal{L}_1$ position becomes out of bounds that wasn't already out of bounds. Every *available* position remains available.) ¶ A *uniquely defining* rule means that the concept is as completely delineated as the prior vocabulary. It introduces no new 'give' into the system that wasn't already there.

Definitions that are conservative and uniquely defining (whether abbreviative or not) can be used, safely, as ways to introduce new concepts, given a background context. They close off no positions that were already open, and they give us new capacities we might otherwise not have had. These are the desiderata used by Nuel Belnap in his response to Prior [1], and I endorse this approach and take myself to be elaborating and developing it.

» «

The inference rules for conjunction, the conditional, and negation, given above are *not* abbreviative definitions. They are also not necessarily conservative, unless we are careful to specify the commitments in our starting language. ¶ Using the conjunction rules alone, we can construct a proof from p, q to p, like so:

$$\dfrac{\dfrac{p \quad q}{p \land q}\,{\land}I}{p}\,{\land}E$$

If our original language $\mathcal{L}_1$ did not tell us that there was a proof from p, q to p, then the addition of these conjunction rules would be non-conservative. ¶ However, our account of proof, motivated in the previous chapter, took it as given that the assumption p on its own, counts not only as a proof from p to p, but also as a proof of p in any context in which p is given, including the context p, q. So, for *us* there remains some chance that these conjunction rules might well be conservative.

If, on the other hand, you wish to explore a more *relevant* notion of consequence, where we admit $p \vdash p$ and exclude $p, q \vdash p$, on the grounds that the q is not used in the derivation of p, then while *these* conjunction rules are unhelpful, other options are open to you [9].

These introduction and elimination rules look rather too much like Prior's rules for tonk for us to be comfortable with them. We will see, though, that with a small tweak, we can see how they are much *closer* to abbreviative definitions than they first appear.

Any rules for $\land$, which introduce it to a vocabulary that has no connectives at all, *cannot* be abbreviative. The context $p, q, \cancel{p \land q}$ cannot be available in the extended language, while $p, \cancel{p \land q}$ and $q, \cancel{p \land q}$ are available. There is no formula in the basic language just containing atomic formulas, that (asserted or denied) clashes with p, q together but does not clash with them individually.

Consider these introduction and elimination rules for the conditional:

$$\frac{\begin{array}{c}[A]^i \\ \vdots \\ B\end{array}}{A \to B}\,{\to}I^i \qquad \frac{A \to B \quad A}{B}\,{\to}E$$

These do not look anything like *abbreviative definitions*. To get closer to the form of an abbreviative definition, let's make the context explicit: The introduction rule says that to derive $A \to B$ from a context $\mathcal{C}$, you can add A to the context and derive B. The elimination rule tells us that if we can derive $A \to B$ in some context, and A in another, we can derive B in a context that shares the assumptions in the two original contexts. We have:

$$\frac{\mathcal{C}, A \succ B}{\mathcal{C} \succ A \to B}\,{\to}I \qquad \frac{\mathcal{C} \succ A \to B \quad \mathcal{C}' \succ A}{\mathcal{C}, \mathcal{C}' \succ B}\,{\to}E$$

If, in the *elimination* rule we restrict our attention to the case where $\mathcal{C}'$ is a context in which A alone is assumed, then the two rules are in fact two directions of the one transition:

$$\frac{\mathcal{C}, A \succ B}{\overline{\mathcal{C} \succ A \to B}}\,{\to}Df$$

This, at least, is a two-way rule, and one that can be treated in just the same way as an abbreviative definition. ¶ If we have a language $\mathcal{L}_1$ *without* the conditional $\to$ in its vocabulary, we can add the conditional, governed by ${\to}Df$ as its defining rule. ¶ This rule is clearly *uniquely defining* (if we happen to apply the defining rule twice, to define two conditionals $\to_1$ and $\to_2$, in any context $\mathcal{C}$, we can prove $A \to_1 B$ if and only if we can prove B from $\mathcal{C}, A$, which holds if and only if we can prove $A \to_2 B$ from $\mathcal{C}$. So, the two conditionals are interchangable.) ¶ That the rule is *conservative* over $\mathcal{L}_1$ requires a rather more work to demonstrate, and the details depend a little on how the starting language $\mathcal{L}_1$ is formulated. The details of exactly what is formally provable are spelled out in Chapter 6 of *Proof, Rules, and Meaning*. Suffice to say, *any* family of rules, for a logical connective, whether ${\to}Df$, or of a form such as these:

A "form such as these" is not too difficult to state: What is below the double line is a general position in which an arbitrary instance of a formula with the newly defined n-ary connective dominant, and above the line are some number of sequents consisting of that position, to which only the constituents of that formula are added.

$$\frac{\mathcal{C}, A \succ \sharp}{\overline{\mathcal{C} \succ \neg A}}\,{\neg}Df \qquad \frac{\mathcal{C} \succ A \quad \mathcal{C} \succ B}{\overline{\mathcal{C} \succ A \land B}}\,{\land}Df \qquad \frac{\mathcal{C}, A \succ C \quad \mathcal{C}, B \succ C}{\overline{\mathcal{C}, A \lor B \succ C}}\,{\lor}Df$$

conservatively extends a base language satisfying the basic structural rules (and perhaps already containing some concepts already given by concepts governed by defining rules from among this family) with the newly introduced concepts. ¶ We can see these rules as showing how to *add* these concepts to our vocabulary, spelling out precisely how to evaluate claims involving the newly defined concepts.

The introduction and elimination rules for *tonk* are not defining rules, and cannot be transformed into such a form, since they do not even conservatively extend a basic language containing only atomic vocabulary.

If we selected either tonk*I* or tonk*E* to be strengthened into an invertible rule, then the result would be safe. However, the first choice would render A tonk B a needlessly complicated way to say A, and the second, a needlessly complicated way to say B.

Viewing the logical concepts as introduced by defining rules, we have a ready answer to how it is that proofs can be gap-free. ¶ Furthermore, the logical concepts used in these proofs are conservative—so they are free to add, in that no pre-existing positions are ruled out by their addition—and uniquely defined. The concepts are as sharply delineated as possible, given the antecedently given field of contents, governed by the structural rules.

» «

Treating these concepts as given by defining rules does not mean that whenever someone uses the everyday concepts, 'and', 'if', 'or', and 'not' they must be given by these rules, or that the everyday concepts agree with the sharply defined concepts $\wedge$, $\rightarrow$, $\vee$ and, $\neg$. ¶ However, insofar as there is unclarity about what we might mean by 'and', 'if', 'or' and 'not', that can always clarify our intentions by using the sharply defined concepts, and settling on using the defining rules as definitional for them.

In what sense do defining rules act as *semantics* for the connectives they govern? In one sense—as giving rules for how to interpret, and to use these concepts—this is completely straightforward. The defining rules are rules for *use* of the governed concepts. ¶ None of this is to do semantics in the model-theoretic sense. We have said nothing about *truth*, or about *truth-in-a-model*.

1.4 POSITIONS & LIMITS / MODELS & TRUTH

There is a natural way to approach models from this proof-first perspective. ¶ Given any available position $\mathcal{C}$, we can say that A is *true-in-$\mathcal{C}$* if $\mathcal{C} \vdash A$, and that A is *false-in-$\mathcal{C}$* if $\mathcal{C}, A \vdash \sharp$. The following facts immediately follow from the defining rules.

- $A \wedge B$ is true-in-$\mathcal{C}$ iff A is true-in-$\mathcal{C}$ and B is true-in-$\mathcal{C}$.
- $A \wedge B$ is false-in-$\mathcal{C}$ *if* A is false-in-$\mathcal{C}$ or B is false-in-$\mathcal{C}$.
- $A \vee B$ is true-in-$\mathcal{C}$ *if* A is true-in-$\mathcal{C}$ or B is true-in-$\mathcal{C}$.
- $A \vee B$ is false-in-$\mathcal{C}$ iff A is false-in-$\mathcal{C}$ and B is false-in-$\mathcal{C}$.
- $A \rightarrow B$ is true-in-$\mathcal{C}$ *if* A is false-in-$\mathcal{C}$ or B is true-in-$\mathcal{C}$.
- $A \rightarrow B$ is false-in-$\mathcal{C}$ iff A is true-in-$\mathcal{C}$ and B is false-in-$\mathcal{C}$.
- $\neg A$ is true-in-$\mathcal{C}$ iff A is false-in-$\mathcal{C}$.
- $\neg A$ is false-in-$\mathcal{C}$ iff A is true-in-$\mathcal{C}$.

Note the *if* in the disjunctive clauses.

Finite available positions are essentially incomplete: They do not settle all issues. ¶ Positions are ordered by extension, adding more assertions or denials. As a position is extended, it makes more things true and more things false. ¶ You can extend the notion of a position to include *infinite* positions. Infinite positions can settle all issues (in a given language $\mathcal{L}$), either positively or negatively. ¶ It is not difficult to show (given the basic structural rules governing positions) that any available finite position $\mathcal{C}$ may be extended, systematically, into a *limit* position $\mathcal{C}^*$, settling all issues, one way or another [8]. ¶ In any limit position $\mathcal{C}^*$, the truth conditions given above hold, except that the disjunctive conditions hold as *biconditionals*, rather than conditionals.

A limit position for $\mathcal{L}$ determines a two-valued boolean valuation on $\mathcal{L}$. If $\mathcal{L}$ is a purely *formal* propositional language, governed by the basic structural rules and the defining rules for the propositional connectives, the limit positions are *exactly* the two-valued boolean valuations, and soundness and completeness for propositional logic follows. ¶ Here, however, connectives are defined in terms of their inference rules. Proofs are primary, and models are secondary.

Limit positions are the inferentialist analogue of models. However, there is more to semantics than truth-in-a-model—We would like to give an account of *truth* [5], and how it relates to truth-in-a-model. What can we say about about truth *per se*? Here, we must step from talking *about* the language we are modelling, to actually *using* it. ¶ Here is how to specify the limit position $\mathcal{C}^*$ which counts as a model of truth. For each sentence A in the language $\mathcal{L}$, if A, then you add A to the context, affirmatively. Otherwise, you add its denial, $\cancel{A}$. ¶ This does not supply an *algorithm* for determining a single position, since we no algorithm for determining truth in $\mathcal{L}$. We have clarified how truth *tout court* is related to truth-in-a-position.

This is the point at which the sentence A is *used*, and not merely mentioned.

If $\mathcal{L}$ contains context-sensitive expressions, here, we assume that the context of use for asserting A when checking for the position is the same context of use *in* the position so evaluated. Matters are more subtle if we wish to include positions containing judgements with different contexts of use, and context-sensitive expressions.

There is more to say about the relationship between inference rules, in general, and truth-conditional theories of meaning, but this is enough for now.

FOR NEXT TIME

ONE QUESTION: A common criticism about inferentalism, is that the semantic relations are language-internal, and are disconnected from *reality*. This criticism is understandable, but mistaken. In the next lecture I will show how an inferentialist semantics can take account of language–world relations.

ONE CHALLENGE: I will also given an account of how inferentialist conditions can shed light on the meanings of modal concepts.

ONE INSIGHT: The answers provided for both issues will involve saying more about the background context governing judgements *as such*. Once that is clarified, the details of the definitions of the quantifiers and modal operators—and their interactions—will be relatively straightforward.

REFERENCES

[1] NUEL BELNAP. "Tonk, Plonk and Plink". *Analysis*, 22:130–134, 1962.

[2] NUEL BELNAP. "Declaratives are not enough". *Philosophical Studies*, 59(1):1–30, 1990.

[3] LEWIS CARROLL. "What the Tortoise Said to Achilles". *Mind*, 4(14):278–280, 1895.

[4] GERHARD GENTZEN. "Untersuchungen über das logische Schließen. I". *Mathematische Zeitschrift*, 39(1):176–210, 1935.

[5] WILFRED HODGES. "Truth in a Structure". *Proceedings of the Aristotelian Society*, 86:135–151, 1986.

[6] DAG PRAWITZ. *Natural Deduction: A Proof Theoretical Study*. Almqvist and Wiksell, Stockholm, 1965.

[7] ARTHUR N. PRIOR. "The Runabout Inference-Ticket". *Analysis*, 21(2):38–39, 1960.

[8] GREG RESTALL. "Truth Values and Proof Theory". *Studia Logica*, 92(2):241–264, 2009.

[9] GREG RESTALL. "Structural Rules in Natural Deduction with Alternatives". *Bulletin of the Section of Logic*, 52(2):109–143, 2023.

[10] GREG RESTALL. "Questions, Justification Requests, Inference, and Definition". *Synthese*, 204(5), 2024.

Greg Restall, *Arché, University of St Andrews* · 2025 Wendy Huang Lectures, National Taiwan University

In the last lecture, I introduced inferentialist semantics, and I showed how you might understand the traditional propositional logical constants as being *definable* by way of invertible inference rules. ¶ These inference rules have distinctive properties. The defining rule for a concept can be read as instructions for how to introduce that concept into a vocabulary that lacks it, in a way that is conservatively extending and uniquely defining. ¶ This gives us a way to make sense of proofs using those concepts as gap-free. The inferential transitions in those proofs are definitionally analytic. ¶ Proofs, in this sense, are at least as semantically significant as models.

In today's lecture, our aim is to extend this account to *quantifiers* and *modal operators*. ¶ As in the last lecture, my exploration is sparked by an insight from Arthur Prior.

Arthur Prior (1914–1969)

> … possible worlds, in the sense of possible states of affairs are not really individuals (just as numbers are not really individuals). To say that a state of affairs obtains is just to say that something is the case; to say that something is a possible state of affairs is just to say that something could be the case; and to say that something is the case 'in' a possible state of affairs is just to say that the thing in question would necessarily be the case if that state of affairs obtained, i.e. if something else were the case … We understand 'truth in states of affairs' because we understand 'necessarily'; not vice versa. [5]

I agree with Prior that possibility and necessity are—in some sense—more conceptually fundamental than possible worlds. ¶ This raises two issues. (1) How *can* we grasp the concepts of possibility and necessity? (2) How is it that those concepts have a structure that makes possible worlds models appropriate for them?

» «

Before we explore how our language might in some sense be understood as describing non-actual possibilities, let's start with the down-to-earth issue of how our thought and talk can be about things in the world we actually inhabit.

2.1 SAYING THINGS ABOUT THINGS

We can make claims *of* things. ¶ I can not only say that Clark Kent can fly. I can say *of* Clark Kent, that *he* can fly.

If I say of that man (who happens to *be* Clark, unbeknownst to me) that *he* can fly, then I am saying, *of Clark Kent* that he can fly.

I will record the distinction between the two kinds of judgements in our formal grammar, distinguishing Fx (the *de re* judgement of x that it's F), which uses a *variable* x, from Ft (the *de dicto* judgement that t is F), where t is a term, such as a name or description, that might feature in our conceptual lexicon. ¶ For the *de re* judgement Fx, the *context of use* picks out the item described. Think of the roles of demonstrative expressions, like 'this', 'it', or 'they'. ¶ For the *de dicto* judgement the rules of the language—concerning names, descriptions, etc.—play a role in selecting the individual, if any, that is named.

One feature relating *variables* and *proof* is SUBSTITUTION. One assumption I will make concerning semantic rules is that the specific identity of the variable in use does not matter to the rules. Any proof featuring '*this*', remains a proof when '*this*' is replaced everywhere by '*that*'. ¶ Formally: Given any proof Π of A from background commitments $\mathcal{C}$, if we replace each free x in the proof by y, the result, Π^x_y is a proof of A^x_y from the background commitments $\mathcal{C}^x_y$. ¶ (All that is required to verify this is to check that each of the *rules* is closed under substitution of one variable for another. This is indeed the case for each of the structural rules and defining rules considered so far, and it is a plausible constraint on any putative lexical rule for an item of our vocabulary.) ¶ The constraint means that these object variables are each *inferentially general*: no rule holds of one variable that does not hold of the others in this class of variables.

DETAILS: Any standard definition of substitution in first-order languages will suffice to make the details precise.

In other words, these variables are appropriately *generic*. If a language contains pronouns (e.g. personal pronouns) that are bound by semantic rules not governing more general pronouns, then the narrower class is not inferentially generic, and the general substitution principle will not hold.

Should this substitution principle hold also for *terms*? If a derivation holds concerning x, would it hold when the term t is substituted for the variable x? ¶ There are reasons to hesitate. Our language might well contain singular terms that do not always take a

REMEMBER: I am interested in logic as it applies to whatever we can think or say. We can make claims in languages with non-denoting terms.

value. There is a number that is $\frac{5}{2}$, but there is no number that is $\frac{1}{0}$. There is no variable that takes $\frac{1}{0}$ as a value, because $\frac{1}{0}$ is not a number, or a thing. '$\frac{1}{0}$' is a non-denoting term.

In a language with terms that might not denote, but in which *variables* may be assumed to always take values—since we interpret Fx as saying *of the item* x that it's F—we need to mark the distinction between a term denoting and a term failing to denote. ¶ Our positions can rule claims in (by assertion) and rule claims out (by denial). It is natural to extend this to say that *terms* can be ruled out, by rejecting them as non-denoting, and ruled *in*, by taking them to denote. ¶ I will represent this by allowing terms (s, t, etc.) and slashed terms ($\not s$, $\not t$, etc.) in our positions, alongside declarative judgements.

Then, we can define an existence predicate, as follows:

$$\frac{\mathcal{C} \succ t}{\mathcal{C} \succ E!t}\; E!Df$$

To say that t exists is to say that the term 't' denotes. ¶ Since variables are assumed to always take values, we impose $\succ x$ as an axiom, for each variable x.

» «

To substitute variables by terms, a weaker principle follows, taking into account the differences between variables and other terms: a free variable may be replaced in a proof by a term, under the extra assumption that the term denotes. This rule is admissible:

$$\frac{\mathcal{C} \succ A}{t, \mathcal{C}^x_t \succ A^x_t}\; Subst^x_t$$

To verify this for a family of rules you must check that each inference rule is closed under substitution of terms for variables, possibly with the added extra premise to the effect that the term in variable denotes. Appeals to the axiom $\succ x$ are replaced by the assumption that t denotes (giving $t \succ t$) and the rest of the proof proceeds as before.

» «

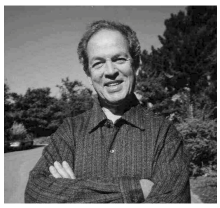
Given the category of singular terms, the class of variables in that category, and substitution, we are now in a position to define the quantifiers, using invertible defining rules. ¶ The universal and existential quantifiers are simply defined:

$$\frac{\mathcal{C} \succ A}{\mathcal{C} \succ \forall x A}\; \forall Df \qquad \frac{\mathcal{C}, A \succ C}{\mathcal{C}, \exists x A \succ C}\; \exists Df$$

In these rules we have the proviso that the variable x bound by the quantifier does not occur free in among the assumptions $\mathcal{C}$ or the conclusion C (in the existential quantifier rule). ¶ To show that *everything* has some feature, we show that x has that feature, while appealing to nothing about x. This is enough to show that *everything* has this feature, since the proof applies to any *thing* at all. ¶ To use the claim that *something* has a given feature as a premise in your reasoning, use the claim that x has this feature, while making no assumptions about what this x is, and not making a claim about x in your conclusion. This reasoning is enough to show your conclusion for it applies, regardless of what item it is that has the feature in question.

Given our assumption that other singular terms may fail to denote, we cannot necessarily conclude $\forall x A \succ A^x_t$. A universally quantified claim only applies to all *things*, and some terms may fail to denote. What does follow is $t, \forall x A \succ A^x_t$, by term substitution. Granted that the term t denotes, if everything has a given feature, t does.

The resulting logic is a simple *free* logic, formulated by Sol Feferman [2, 7]. ¶ If we add the extra assumption that terms denote, this is no more and no less than classical first-order predicate logic.



Solomon Feferman (1928–2016)

» «

It is reasonably straightforward that the quantifiers, so defined, satisfy the conservative extension and uniqueness criteria. ¶ For CONSERVATIVE EXTENSION, the argument proceeds in the same way as for the propositional constants. (We pass from the invertible defining rule to a pair of left and right sequent rules, and show that the *Cut* rule can be eliminated, and since the remaining rules satisfy the subformula property, if anything can be proved, in the reformulated system, it has a proof that uses only the logical concepts occurring in the item to be proved. So, adding new connectives or quantifiers does not render anything in the old vocabulary provable that was unprovable before that addition.) ¶ For UNIQUENESS, since the quantifiers are defined by an invertible rule, we can pass from $\forall_1 x A$ to $\forall_2 x A$ through the old language, as before.

I have left out details concerning the behaviour of predicates and function symbols. To get Feferman's negative free logic, we add rules of the form $Ft \succ t$ and $ft \succ t$, constraining predicates to apply to terms only when they denote, and functions having defined values only when their inputs are defined.

However, it is worth reflecting on how uniqueness is obtained, since it is well known that there are multi-sorted first-order languages. It is straightforward to have a language with a quantifier that ranges over *objects* (say) and another that ranges over *locations* (say), where predicates are sorted (some apply to objects or relate objects to objects; others apply to or relate places; and yet others relate objects and places). We have object variables and location variables, and as a result we have *distinct* object quantifiers and place quantifiers. ¶ None of this contradicts the uniqueness result given here, since the uniqueness that we have shown depends on the grammatical categorisation of the language under examination. In the derivation showing that $\forall_1$ and $\forall_2$ are equivalent quantifiers, we assume that 1-variables are 2-variables, and *vice versa*. ¶ The meaning of the quantifier expressions given inferentially depends not only on the structural rules governing the field of issues being reasoned about (as we saw in the first lecture), but they *also* require fixing on a grammatical category of singular terms.

This will become a live issue in the next lecture, where we consider what is involved in quantifying over *possible worlds*.

» «

I have given inference rules for the quantifiers that agree with the standard truth conditions. In none of this discussion have I given a semantics featuring a domain of objects over which the quantifiers range. ¶ Since the inference rules use a notion of substitution, and since substitution in formulas can be defined purely *syntactically*, you might think that this is a substitutional account of the quantifiers [1, 4], according to which a quantified expression $\forall x Fx$ [$\exists x Fx$] is true if and only if every [some] substitution instance of Fx is true. ¶ This is a mistake, and not only because we have allowed for non-denoting terms. Even if we imposed the condition that every term denotes, the truth of every instance Ft, Ft', Fx, Fy, etc., would not be enough to guarantee the truth of $\forall x Fx$. ¶ There is no way to prove the conclusion $\forall x Fx$ from the premises Ft, Ft', Fx, Fy, etc., no matter *how* many terms of the language are included. So, whatever $\forall x Fx$ says, it must say something more than what is said by the instances taken together. This inferentialist semantics is not substitutional in *that* sense.

Modulo a choice for whether to allow non-denoting terms.

Take a model in which there are at least two objects, one of which has feature F and another of which does not, and let's interpret our language so that every term denotes the F-object. Furthermore, assign the value of each variable, *that* object, too. Here, the premises are satisfied, but the conclusion $\forall x Fx$ is not.

What this means for models, domains, and how we *should* interpret the quantifiers, I will leave for Section 4. ¶ Before we get there, let's consider modality.

2.2 SUPPOSING THINGS WERE OTHERWISE

To add quantifiers, we exploited the structure of a language, isolating a class of singular terms. For modal logic, instead, we will return to the kinds of speech acts we can use in our thought and talk. ¶ When we *modalise*—considering matters concerning what is possible and what is necessary—we consider not only what *is* the case, but what *might be*. We do not just assert and deny and ask questions. We *suppose*.

We have already seen supposition, when considering the defining rules for the material conditional and for negation. To prove $A \to B$, we suppose A and derive B. To prove $\neg A$, we suppose A and reduce the result to absurdity. ¶ In these cases, supposition is simple. To suppose A we temporarily add A to the current discursive context, to see what additional consequences we might draw out. (If the current context already rules A out, then this is good news when the aim is to prove $\neg A$.)

Recall, 'absurdity' here is marked in proof by '$\sharp$' the sign that we have reached a dead-end, a self-undermining position.

Not all acts of supposition merely add the supposed item to the current context. ¶ When we plan for the future, it is common for us to wonder what might happen, and to take various outcomes into account. We might plan for rainy weather tomorrow and for sunshine. ¶ We might (for the moment) suppose that it *does* rain, and consider this alternative, and then suppose that it *doesn't* rain, and consider what we might do under those circumstances. ¶ That is not very different to ordinary supposition, because we typically do not have very reliable information about the future. However, we can also make suppositions in this way *about the past*. We can reflect, in just the same way, about what would have happened *had it not rained*, or *had it rained*.

To suppose *it hadn't rained yesterday* is to do something quite distinctive. It's not just to add the claim 'it didn't rain yesterday' to my current commitments (or to the common ground in the conversation). That would be a short step to absurdity, in a context where we grant that it *did* rain yesterday. ¶ Instead, this kind of *modal* supposition asks us to

There is also a characteristically *epistemic* kind of supposition, arising when we attempt to rationally manage disagreement. You and I might disagree over p, and I might say: suppose you're right, and that it's actually the case that p . . . Here, I do not suppose p counterfactually in the subjunctive sense important for planning. Here, I consider p as part of an alternative account of *how things actually are* [3, 6]. We will not consider this epistemic supposition and the associated modality here, but the techniques under consideration apply equally well to epistemic modality.

reason with this supposed content in a different way. It is isolated from our current commitments (about *how things actually went*), but we reason in the same way that we would when we flatly reason about how things actually turn out. ¶ I will not focus on the detail of how to counterfactually reason about local circumstances. My focus here is on the idealised modal operators of *possibility* (◊) and *necessity* (□) broadly understood.

In particular, I won't concern myself about what from our actual commitments we import to the other context, when we suppose p in that local counterfactual sense, of changing the current circumstances only so far as to admit p.

» «

The key notion for us is that in modal reasoning we can understand 'supposing things were otherwise' and 'considering alternative possibilities' in a straightforward way. We apply our thought and talk not just to a flat position consisting of assertions and denials reflecting our current commitments, but to a richer structure, which might look like this:

$$[_{@}\ p, q, \cancel{r} \mid p, s, \cancel{r} \mid r, \cancel{p}, \cancel{q}\,]$$

Such a richer position represents the commitment of someone who has granted p and q and denied r (that is their '*actual*' commitment, as marked by the '@') and who has admitted as *possible* two other alternative outcomes: p and s without r or r without p and q. ¶ Such enriched positions, with more than one ZONE, arise when we engage in counterfactual reasoning, and plan in the face of uncertainty. ¶ There is no conflict between asserting p in one zone and denying it in another, because we are considering alternatives. ¶ In such a context, the semantics of modal operators comes to light. To say that something is necessary is to preserve that commitment across alternatives. So, a position like this is defective:

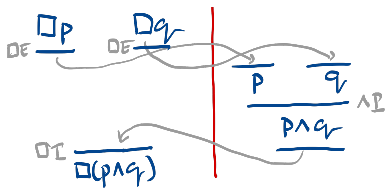
$$[\cdots \mid \Box A \cdots \mid \cancel{A} \cdots \mid \cdots]$$

To grant (somewhere) that A is *necessary* and to deny A (elsewhere) is to contradict yourself. ¶ Similarly, this position is also out of bounds:

At least, to grant □A in some zone and to deny A in another is to contradict yourself if the notion of necessity in question encompasses the kinds of alternatives represented by the zones in our position.

$$[\cdots \mid \cancel{\Diamond A} \cdots \mid A \cdots \mid \cdots]$$

This motivates modal reasoning, which could be represented like this:



This reasoning presumes that the semantic rules for conjunction apply in *other* zones, as much as they do in our 'home' zone. That is the constraint that we will impose throughout. We take our definitions to apply to the language as we use it *across supposition boundaries*. (How *else* could we employ our concepts in subjunctive reasoning?)

The left side of the line corresponds to our starting zone, and the right side, the other zone, under the supposition that things had gone otherwise. ¶ This graphical representation is hard to typeset. Instead, we can distinguish different zones with tags. Here, 'a' labels commitments in the starting zone, while 'b' labels those in the other zone.

$$\dfrac{\dfrac{\dfrac{\Box p \cdot a}{p \cdot b}\,\Box E \qquad \dfrac{\Box q \cdot a}{q \cdot b}\,\Box E}{p \wedge q \cdot b}\,\wedge I}{\Box(p \wedge q) \cdot a}\,\Box I$$

The defining rules for the modal operators then take a form reminiscent of the quantifier rules, but which arise out of the richer structure of the zones in our positions, rather than quantifying over *objects* to which we might be able to *refer*.

$$\dfrac{\mathcal{C} \succ A \cdot i}{\mathcal{C} \succ \Box A \cdot j}\,\Box Df \qquad \dfrac{\mathcal{C}, A \cdot i \succ C}{\mathcal{C}, \Diamond A \cdot j \succ C}\,\Diamond Df$$

An *actuality* operator may be defined with a rule defining @A · a to have the same effect as A · @. Granting @A in any zone of the discourse has the same effect as granting A in the *actual* zone. This, in combination with the rules for □ and ◊, is a simple family of rules for the standard logic S5@.

Here, the important side condition in this rule is that the zone tag i in the premise is absent from the assumptions $\mathcal{C}$ (and the conclusion C in the ◊ rule). ¶ Our demonstration concerning A is *arbitrary*, no matter what else we have supposed in that counterfactual zone. So, whatever we have concluded applies *generally*.

» «

These rules are CONSERVATIVELY EXTENDING and UNIQUELY DEFINING, for the same reasons as for the quantifiers, except that this reasoning depends on supposition, rather than *substitution*. ¶ The uniqueness result requires coordination on what counts as the relevant kind of *modal supposition*.

2.3 BUT WHAT IF *THIS* THING HADN'T EXISTED?

Now let's combine the quantifier rules and the modal rules. ¶ Here is a natural proof:

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{[\forall x\Box Fx \cdot a]^1}{\Box Fx \cdot a}\;\forall E}{Fx \cdot b}\;\Box E}{\forall x Fx \cdot b}\;\forall I}{\Box\forall x Fx \cdot a}\;\Box I}{\forall x\Box Fx \to \Box\forall x Fx \cdot a}\;{\to}I$$


Ruth Barcan Marcus (1921–2012)

We can prove the infamous BARCAN FORMULAS, named after Ruth Barcan Marcus, a pioneer of the study of first-order modal logic. ¶ If everything is necessarily F, then it is necessary that everything is F.

» «

There are reasons to resist the Barcan formulas. It is natural to think that there could have been things that don't actually exist, and that some of what actually exists could have not existed. ¶ If everything that *happens* to exist is necessarily an F, it follows only that it is necessary that everything (that happens to have existed) is also an F only if there could be no *other* things than those things that happen to exist. ¶ Our quantifier rules, as given in Section 2.1, make no distinction between what exists and what doesn't, because our background context makes the assumption that variables always have values, and so, they always (and in any context) denote existing objects.

This assumption can be rejected, and it *should* be rejected if you take it to be coherent to suppose, of a given item, that *it* might not have existed. ¶ This points to an ambiguity between a variable *being defined* and *having a value*. When we take a given item x (say *me*) and suppose x *had not existed*, then the variable x is *is defined*, but we are considering a circumstance where in which the value of that variable does not exist. ¶ This motivates revising the variable rule. Instead of requiring that variables have values that present in *every* zone, we require only that their values be present in *some* zone.

$$\frac{x \cdot i, \mathcal{C} \succ C}{\mathcal{C} \succ C}\;\Diamond Var$$

(Here, the zone i must not be used elsewhere in $\mathcal{C}$ and C.) ¶ With this in place, the quantifier rules come in two forms: one is existentially committing in the current zone, and the other ranges across all zones in scope.

$$\frac{\mathcal{C}, x \cdot j \succ A \cdot j}{\mathcal{C} \succ \forall x A \cdot j}\;\forall Df \qquad \frac{\mathcal{C}, x \cdot j, A \cdot j \succ C}{\mathcal{C}, \exists x A \cdot j \succ C}\;\exists Df \qquad \frac{\mathcal{C} \succ A \cdot j}{\mathcal{C} \succ \forall^{\Diamond} x A \cdot j}\;\forall^{\Diamond} Df \qquad \frac{\mathcal{C}, A \cdot j \succ C}{\mathcal{C}, \exists^{\Diamond} x A \cdot j \succ C}\;\exists^{\Diamond} Df$$

In these rules we have the proviso that the variable x bound by the quantifier does not occur free in among the assumptions $\mathcal{C}$ or the conclusion C.

With the existentially committing rules for ∀ and ∃, the Barcan proofs fail, without the principle to the effect that something existing *there* also exists *here*. We start …

$$\cfrac{\cfrac{x \cdot a \qquad \forall x\Box Fx \cdot a}{\Box Fx \cdot a}\;\forall E}{Fx \cdot b}\;\Box E$$

… but stop there. We cannot introduce the universal quantifier unless we can derive Fx at b under the supposition that x's value is present *at* b. Unless we can derive x · a from $x \cdot b$, we are stuck. ¶ One way to verify that we *cannot* derive $\Box\forall x Fx$ from $\forall x \Box Fx$, using these rules, is to systematically reason about what can be derived. ¶ Instead, we will consider what this perspective can tell us about *models*, and use this to give counterexamples to invalid reasoning.

2.4 PUSHING THINGS TO THE LIMIT

In Lecture 1, we saw how *limit* positions—the available positions that decide every issue in a given language—provide another way to think of traditional two-valued models. ¶ Given a purely propositional language $\mathcal{L}$, with no quantifiers or modal operators, these positions settle all $\mathcal{L}$. ¶ Quantifiers and modal operators complicate this picture.

Start with *quantifiers*: Consider a tiny language $\mathcal{L}^{F}_{\exists}$ with one unary predicate F, no connectives, a family of variables and singular terms, and the existential quantifier, governed by the defining rule $\exists Df$. ¶ There is no contradiction in affirming $\exists x Fx$ while denying Ft for every term t in the language (including variables). A position denying every formula of the form Ft and affirming $\exists x Fx$ is available, and can be extended into a partition of all of $\mathcal{L}^{F}_{\exists}$. ¶ There is an issue left unsettled by this position: we have affirmed $\exists x Fx$, but have no insight into *which* item has feature F.

This is a fact concerning the inference rules for $\mathcal{L}^{F}_{\exists}$. There is no way to derive a contradiction from $\exists x Fx$ and any number of denials of the form $\cancel{Ft}$.

Let's call a position $\exists$-WITNESSED if whenever it includes $\exists x A$, it also includes A^x_y for some variable y. (We say, *of some item* y that it has the relevant feature.)

The defining rule for $\exists$ ensures that any available finite position including $\exists x A$, is extended by an available position that contains A^x_y for some fresh variable y. ¶ (Why is this? Suppose that our finite position contains $\exists x A$. Choose a variable y not already present in the position. Given any proof leading from premises in that position together with A^y_x to a contradiction, since y is absent from the position, we can use $\exists Df$ to derive the contradiction from $\exists x A$ alone. Since the original position is available, the extended position is, too.) ¶ This ensures that any available (small) position may be extended to a limit position that is also $\exists$-*witnessed*.

The finiteness condition ensures that we can choose a variable that is, as yet, unused by the position. This can be weakened in natural ways, but those details are not necessary for our purposes.

What goes for $\exists$ goes also for $\forall$. I leave those details for you to consider.

» «

The modal operators raise parallel issues. Given a position in which I affirm $\Diamond A$ (in zone i), this raises the question: *How* could A have been the case? To appropriately *settle* the issue, our position also involve a zone in which A is affirmed. That will show how this issue can be answered. (Perhaps at the cost of raising further issues!) ¶ Call positions which contain, for each $\Diamond A \cdot i$, some $A \cdot j$, $\Diamond$-WITNESSED. ¶ Just as with the $\exists$ rule, that this constraint can be met arises from the defining rule $\Diamond Df$. Any available position containing $\Diamond A \cdot i$ may be extended to contain $A \cdot j$ for a fresh zone j.

If I say it's possible that $2+2=5$, you're within your rights to ask me *how*, and if you suppose that $2+2=5$ and reduce that (counterfactual) supposition to absurdity, using agreed premises and principles, you've ruled out my claim that it's possible.

What goes for $\Diamond$ goes also for $\Box$. I leave those details for you to consider.

Witnessed limit positions in a language $\mathcal{L}$ settle every $\mathcal{L}$-issue in this extended sense.

» «

Unsurprisingly, witnessed limit positions describe Kripke models. ¶ Each *zone* describes what is true at a possible world. The terms ruled *in* at a zone are the inner domain of objects *existing* at that world, while the remaining terms that are ruled in at *some* world are the outer domain of *possibilia*. ¶ The usual truth conditions for a varying domain quantified S5 (with actuality) are satisfied, given the defining rules for the quantifiers, connectives and modal operators, and the witnessing conditions governing the construction of our position.

For the technical details, see my draft "Modal Logic and Contingent Existence (Generality and Existence 2)" [8].

Here is a sketch of how such positions are constructed. Start here, attempting to refute a Barcan formula, affirming $\Diamond\exists x Fx$ and denying $\exists x \Diamond Fx$.

$$[_{@}\Diamond\exists x Fx, \cancel{\exists x\Diamond Fx}]$$

For the witnessing condition for $\Diamond$ we must add a zone affirming $\exists x Fx$. We have:

$$[_{@}\Diamond\exists x Fx, \cancel{\exists x\Diamond Fx} \mid \exists x Fx]$$

We must add a witness for $\exists x Fx$, which is also present at that zone.

$$[_{@}\Diamond\exists x Fx, \cancel{\exists x\Diamond Fx} \mid x, \exists x Fx, Fx]$$

Since Fx is affirmed in the second zone, $\Diamond Fx$ is undeniable in each zone (given $\Diamond Df$), so as we fill out the position, $\Diamond Fx$ must be settled affirmatively in each zone.

$$[_{@}\Diamond\exists xFx, \cancel{\exists x\Diamond Fx}, \Diamond Fx \mid x, \exists xFx, Fx, \Diamond Fx]$$

Now, since $\exists x\Diamond Fx$ is *denied* but $\Diamond Fx$ is affirmed in the first zone, the object x must fail to be present there.

$$[_{@}\Diamond\exists xFx, \cancel{\exists x\Diamond Fx}, \Diamond Fx, \cancel{x} \mid x, \exists xFx, Fx, \Diamond Fx]$$

If the predicate F expresses an existence entailing property (as all predicates do in Feferman's logic) then Fx must also be rejected in our starting position:

$$[_{@}\Diamond\exists xFx, \cancel{\exists x\Diamond Fx}, \Diamond Fx, \cancel{Fx}, \cancel{x} \mid x, \exists xFx, Fx, \Diamond Fx] \qquad (\mathfrak{P})$$

This position describes a tiny model, settling every issue concerning the single object x and the predicate F. It provides a counterexample to the Barcan formula.

» «

This argument provides an solution to Prior's modal semantics puzzle. Possible worlds models manage to model modal vocabulary when that vocabulary is governed by defining rules, since witnessed limit positions describe these models.

» «

ISSUE 1: What about the *ontological* significance of possibilia and of possible worlds?

Witnessed limit positions are *abstracta*. They are *constructions* arising out of our linguistic practice. However, saying that these models are *abstracta*, and therefore, that there are no interesting issues in the ontological commitment undertaken when adopting those models is not addressing the important issues involved in modal semantics.

The *vital* issue concerns what it is to *take* such a position. To *endorse* a position is to take it that what is in the zone marked @ holds, while what is in the other zones is *possible*. So, according to ($\mathfrak{P}$), $\Diamond\exists xFx$ is true, and it's jointly possible (1) that there exists some item, x, and (2) it is F. From this it follows in *that* scenario that $\exists xFx$ and $\Diamond Fx$. Also, according to ($\mathfrak{P}$), the item featuring in that other possibility is actually existent.

This is the kind of ideological and ontological commitment involved in taking these models to describe *how things are*. They systematically spell out the commitments made when using modal and quantificational vocabulary. ¶ What more do you want from your models than that?

» «

ISSUE 2: The models given by constructing witnessed limit positions are a *scale models*. They need not be one-to-one models of the universe. ¶ Suppose $\mathcal{L}$ is countably infinite. You can construct a witnessed limit position with only countably many items in the domain, no matter what your starting position is like. ¶ You could start with a standard mathematical theory, like ZFC, according to which there are uncountable sets. ¶ Regardless, the limit position constructed will be countably infinite, from the outside. We have countable models of ZFC. This is familiar to mathematicians.

Other models are possible, of course. But they aren't made by *this* construction.

Does this mean that $\forall$ doesn't mean *all*, that one cannot quantify over *absolutely everything*? No. A model, as a set-theoretic object, is a *model*, of the language, it is not the language as it is *actually used*. The *reality* is what is described by the vocabulary in use. ¶ You are genuinely *saying* that everything has feature F when you say $\forall xFx$, even if the *models* representing that statement are small and do not exhaust everything. ¶ As conceptual capacities expand, and our language is extended, our original commitment to $\forall xFx$ will, if it is retained, be about the things we talk about *then*, too. ¶ As we saw before, the semantics of the quantifiers are not substitutional in the flat sense, and there is no bar to '$\forall$' meaning *all*, even if every *model* of this vocabulary is small.

Similarly, if models are *classes*. That merely punts the issue up to another level. If a model is a *thing*, it is extremely difficult (if not inconsistent) for a one-to-one scale model of the universe to contain itself.

Similarly, there is no requirement here that an adequate $\mathcal{L}$-model of a modal language must exhaust modal space. An adequate model for an extended language $\mathcal{L}'$ may have more worlds. (Zones may be $\mathcal{L}$-indistinguishable without being $\mathcal{L}'$-indistinguishable, in the sense that a complete $\mathcal{L}$-zone might be expanded with $\mathcal{L}'$ vocabulary coherently in two different ways.) Again, that is unproblematic. The models of a modal language are tractable representations of what is possible and what is necessary, describing 'worlds' only as far as needed for the purposes of that spelling out commitments expressible in that language. ¶ So, there is no problem in using a model of modal vocabulary, without having to address thorny questions settling how many objects or worlds there in fact are. A model can be adequate for representing statements expressible in that language without that model settling once-and-for-all difficult questions how many objects or worlds there are. *That* is not what models are for.

FOR NEXT TIME

We will take this perspective on models for modal vocabulary, and see how this can give a new angle on the value of truth-conditional semantics for natural languages.

REFERENCES

[1] J. MICHAEL DUNN AND NUEL BELNAP. "The Substitution Interpretation of the Quantifiers". *Noûs*, 2(2):177–185, 1968.

[2] SOLOMON FEFERMAN. "Definedness". *Erkenntnis*, 43(3):295–320, 1995.

[3] MARK LANCE AND W. HEATH WHITE. "Stereoscopic Vision: Persons, Freedom, and Two Spaces of Material Inference". *Philosophers' Imprint*, 7(4):1–21, 2007.

[4] RUTH BARCAN MARCUS. "Interpreting Quantification". *Inquiry*, 51:252–259, 1962.

[5] ARTHUR N. PRIOR AND KIT FINE. *Worlds, Times and Selves*. Duckworth, 1977.

[6] GREG RESTALL. "A Cut-Free Sequent System for Two-Dimensional Modal Logic, and why it matters". *Annals of Pure and Applied Logic*, 163(11):1611–1623, 2012.

[7] GREG RESTALL. "Generality and Existence 1: Quantification and Free Logic". *Review of Symbolic Logic*, 12:1–29, 2019.

[8] GREG RESTALL. "Modal Logic and Contingent Existence (Generality and Existence 2)". Draft available at `https://consequently.org/writing/mlce-ge2/`, 2025.

In my first lecture, I introduced INFERENTIALIST SEMANTICS, giving an account of how you can understand propositional logical vocabulary as *definable* using invertible rules of inference. This gives you a ready explanation of how it is that a proof can be completely gap-free. To adopt $\rightarrow Df$ as definitional of the material notion of implication is to take it that to show $A \rightarrow B$, it suffices to show B while taking A for granted, and conversely, if we have granted $A \rightarrow B$ and granted A, then B must follow. To ask for anything more is to fail to take '$\rightarrow$' to be defined by $\rightarrow Df$. The defining rules, together with the fundamental structural rules governing proof *as such* are the only transitions in our gap-free proofs. Since these defining rules conservatively extend those background structural rules, they are safe to adopt.

$$\frac{\mathcal{C}, A \succ B}{\mathcal{C} \succ A \rightarrow B}\ \rightarrow Df$$

In the second lecture, I showed how this account generalises to quantifiers and to modal operators, with one difference. The defining rules for these concepts appeal to distinctive features of our linguistic and conceptual practices. For the first order quantifiers, our language must involve a category of singular terms, satisfying constraints on substitution of one term for another.

$$\frac{\mathcal{C} \succ A}{\mathcal{C} \succ \forall x A}\ \forall Df \quad (x \text{ is not free in } \mathcal{C}.)$$

The defining rules for the modal operators involve expanding our background structural account of proof to encompass not only acts of assertion, denial, inference and purely material supposition (consisting of temporarily granting something 'for the sake of the argument', by adding it to the common ground, leaving the rest unchanged), to include properly *modal* supposition. To modally suppose A is to consider what would be the case *had* A obtained. This is, of course, compatible with denying A, and it does not involve withdrawing your denial of A. To modally suppose A is to add A to the conversational context in a *new zone*. We use our reasoning capacities with this considered content in its zone, isolated from our commitments concerning how things are. The isolation is not total, because cross-zone commerce is underwritten by modal operators. To say take it that A is necessary is to take it that A not only holds, but *would* hold had things gone otherwise. Tagging zones with labels, there is no clash between asserting A at a and denying it at another zone b, but there *is* a clash between asserting $\Box A$ at a and denying A at b. Denying A at b is, implicitly, denying that A is necessary.

$$\frac{\mathcal{C} \succ A \cdot i}{\mathcal{C} \succ \Box A \cdot j}\ \forall Df \quad (i \text{ is not present in } \mathcal{C}.)$$

This modal logic contains a varying domain inner *existence-entailing* quantifiers and constant domain outer *possibilist* quantifiers, given the discipline governing variables discussed in the previous lecture.

Combining defining rules for the first-order quantifiers and modal operators gives an inferentialist semantics for first-order quantified modal logic, where the inference rules governing the logical connectives describe what we *do* to use the concepts in making assertions, denials, inferences, and suppositions, i.e., to use them in our talk and thought. ¶ Adopting an inferentialist semantics need not involve *rejecting* traditional truth-conditional model theory for that vocabulary. I showed how familiar models for first-order modal logics arise as the limit of a process of settling issues in the structured *positions* that feature in inference, and that the model-theoretic 'semantics' for modal logic can be defended on inferentialist lines. This is one way to vindicate Arthur Prior's view that we understand possible worlds because we first understand the notions of possibility and necessity. This inferentialist semantics shows how to acquire the concepts of possibility and necessity in such a way that possible worlds talk is a model of its logical structure.

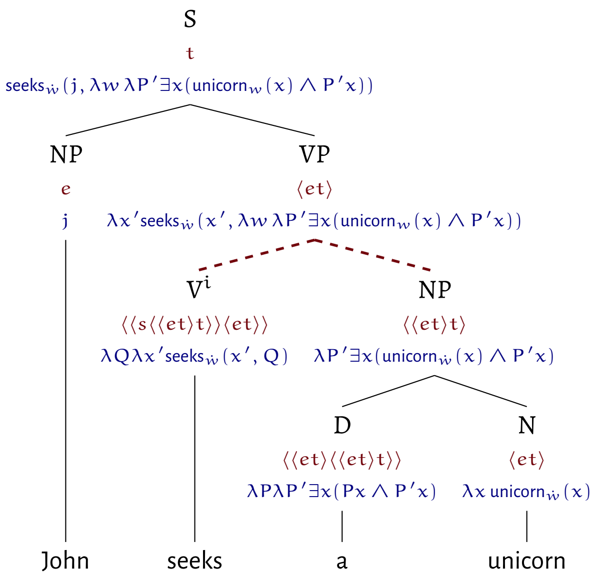
One motivation for this line of inquiry this challenge from Timothy Williamson. Speaking of Robert Brandom, he says his "inferentialism has remained at an even more programmatic stage than Dummett's, lacking an equivalent of Dummett's connection with technical developments in proof theory by Dag Prawitz and others. As a result, inferentialism has been far less fruitful than referentialism for linguistics. In that crude sense, referentialism beats inferentialism by pragmatic standards" [11, p. 34]. While I do not so much care about what perspective 'beats' another, Williamson does prompt the question of how inferentialism bears on model-theoretic semantics and its 'referentialist' commitments.

3.1 THERE IS MORE TO LANGUAGE THAN LOGIC

This inferentialist semantics has many virtues. However, I have been playing on the logicians' home turf, focussing on logical connectives, quantifiers, and modal operators. There is more to language than this tiny fragment. While it is not my job to give a comprehensive meaning theory for every concept that we can use in our thought and talk, I would do well to explain how the inferentialist 'semantics' I offered for the logical concepts might relate to the kinds of semantic theories offered by linguists and philosophers of language.

Semantics is an incredibly rich and diverse field, and I cannot do justice to it in a single lecture [6]. I will focus on the *truth-conditional* semantics, as pioneered by Richard Montague [10], David Lewis [5] and Barbara Partee [7, 9], in the 1970s. Modal model theory is central to this enterprise. Semantic values are supplied for natural language lexical items, in the vocabulary of an *intensional type theory*. Lexical items of different parts of speech are interpreted as different *types*, which compose in a compositional manner, to provide an interpretation for a complex construction out of the values of its constituents. The theory is *intensional* because the basic semantic values are not merely composed out of this-world referents of singular terms, and extensions of predicates. While we can understand what it is for John to *greet* a unicorn in terms of the relationship between John and any individual unicorns, what it is for John to *seek* a unicorn is a more subtle matter. In an intensional type theory, an intensional transitive verb expects the *value* of a noun phrase as one input (the *seeker*), but takes the *intension* of a noun phrase (what it is that is sought) as the other. After all, one can count as seeking a unicorn without there being any unicorns there to be found.

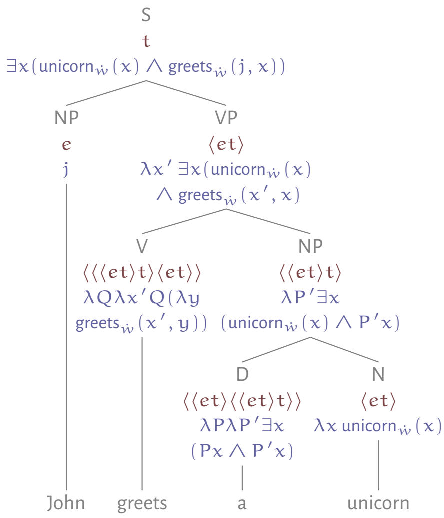
In what follows, I will consider a simple, general framework for intensional type theory, the two-sorted type theory Ty_2 of Daniel Gallin [3]. It is a natural extension of a two-sorted first-order logic, with one sort e for *entities* and another sort s for *states*. These basic sorts are also the basic *types* of the vocabulary. *Sentences* in the language have type t (on the intended semantics, each sentence has a *truth value*). ¶ For any type α and β there is a *functional* type, $\langle\alpha\beta\rangle$ which can be thought of as functions from items of type α to items of type β. A one-place predicate, then, has type $\langle et\rangle$ since when combined with an item of type e (a singular term) as input and returns a sentence as an output. At the level of *values*, a term of type $\langle et\rangle$ can be interpreted as a function from objects to truth values. A term of type $\langle s\alpha\rangle$ is a function from *worlds* to values of type α. ¶ Here is a Ty_2 derivation of a semantic value for the expression 'John seeks a unicorn':



Here, the NP 'a unicorn' has a complex semantic value given by applying the determiner 'a' to the unary predicate 'unicorn', where the latter is interpreted at $\dot{w}$, the world of evaluation. Since 'seeks' is an *intensional* verb, the input NP value is *intensionalised* by binding the world of evaluation. As a result, there is no requirement that John's quest be backed up by any existing unicorn. ¶ We do not need to pursue the details of this compositional meaning theory. It is enough to notice that it is denotational semantics, and is committed to a basic framework of possible worlds, entities, and truth values, and a hierarchy of grammatical and semantic types corresponding to functions between those different categories. ¶ It has been remarkably successful in analyses of complex compositional behaviour in natural languages. My task is to situate inferentialist semantics with respect to this tradition.

Barbara Partee (1940–)

Here is a Ty_2 derivation for 'John greets a unicorn'. Notice that for it to be true that John greets a unicorn, there must *be* a unicorn that John greets.



Elizabeth Coppock and Lucas Champollion's textbook is a particularly comprehensive and accessible introduction to intensional type theory in general, and Ty_2 in particular [2].

In selecting truth-conditional semantics for my target, I do not mean to ignore or to downplay the many *other* issues in semantics, concerning how it is that the basic lexical items get their semantic values, how to understand context sensitivity, ambiguity, and much else besides [6]. Those are important issues, but they are orthogonal to my central aim concerning the relationship between proof theory and model theory, and what this might say, more generally, about the relationship between representational semantics and inferentialist semantics. ¶ My focus here is to consider the power of truth-conditional semantics, to revisit what kinds of commitments are involved in adopting it, and what it can do for us, in the light of inferentialist semantics for quantified modal logic. Exactly what kind of commitment is involved in the kind of 'world' talk that is in use by linguists.

For example, when John seeks a *unicorn*, is he seeking a mythical horned horselike creature, or is he after a privately held startup with an obscenely high valuation? Is the difference between the two kind of 'unicorn' a matter of two different lexical entries, or contextual variation concerning the one vocabulary item?

Linguists have noticed that their commitments concerning the entities and worlds of their semantic theories are not quite the same as the commitments of metaphysicians. Here is Barbara Partee, writing in the late 1980s, during the heyday of the discussion of the metaphysics of possible worlds.

> A non-absolutist picture seems to fit linguistic semantics better than an absolutist one, where by absolutist I mean the position that there is one single maximal set (or class, if it's too big to be a set) of possible worlds. If a philosopher could find arguments that in the best metaphysical theory there is indeed a maximal set, I suspect that would for the linguist be further confirmation that his enterprise is not metaphysics, and I would doubt that such a maximal set would ever figure in a natural language semantics. As various people have noted, *possible worlds* are really not so different in this respect from *entities*: every model-theoretic semantic theory I'm familiar with takes entities to be among the primitives—but puzzles about the identity conditions of individuals and about whether there is a maximal set of all of them are just as problematic, and it is just as questionable whether semantic theory has to depend on settling such questions. [. . .] it is the structure provided by the possible worlds theory that does the work, not the choice of particular possible worlds, if the latter makes sense at all. [8, p. 118]

How should we understand the semanticists commitment to a domain of possible worlds and to a domain of entities in their semantic theories? How does the *structure* provided by possible worlds theory do the work, rather than the choice of 'particular possible worlds', as Partee puts it? ¶ My own theoretical orientation when answering this question is owed to my second inspiration in these talks, Nuel Belnap. He writes:

Nuel Belnap (1930–2024)

> … in the tolerant spirit of Carnap, we believe that one is likely to want a *variety* of complementary (noncompeting) pre-semantic analyses—and most especially, a variety of pre-semantic treatments of one and the same 'language.' One does not have to 'believe in alternative logics' to repudiate the sort of absolutism that comes not from logic itself, but from narrow-gauge metaphysics or epistemology … although Carnap's beneficent influence is legendary, it seems worth repeating the lesson: There can and should be multiple useful, productive, insightful and pertinent analyses of the *same* target. Pre-semantics therefore emphasizes the usefulness of thinking in terms of a *variety* of pre-semantic systems. [1, p. 1]

Belnap's salutary advice is ignored, far too often. Of *course* we can benefit from a number of non-competing analyses of the one target. Different accounts of the one phenomenon will foreground distinct features. One can be quite satisfied with the traditional truth-conditional semantics for intensional type theories, while being curious about how that structure is to be understood, how it is grounded in reality, and connected to our practice. ¶ I will offer an inferentialist *pre*-semantics for this truth-conditional semantic framework. The aim is to vindicate Barbara Partee's judgement, to the effect that possible worlds models serve as an incredibly useful structuring device, but it need not follow that linguists (or any of us) must fix on some antecedently given, metaphysically privileged class of possible worlds. The inferentialist pre-semantics will help us give a fresh account of the function of 'possible world' talk in truth-conditional semantics. ¶ Furthermore, this inferentialist pre-semantics will have as a corollary, an account of how the *intension* of a natural language expression (given as a function of type $\langle s\alpha \rangle$ in the truth-conditional semantics) might correspond to something that a language user might have the capacity to *grasp*, to a greater or lesser extent, bringing *representational* and *use-based* semantic accounts into a common framework. ¶ To bring our reflections full-circle, I will then show how the defining rules for the logical expressions count as giving the meanings of those expressions, understood both in terms of truth conditions and rules for use.

3.2 TAKING A GOD'S EYE VIEW

First, recall that *positions* in our inferentialist account of the kind of language use salient for modalising are structured into distinct zones. One zone distinguished as the 'actual' zone, in that the commitments recorded there are taken to apply to how things actually are. The other zones record claims that are endorsed as *possible*, which are considered as *alternatives*. A position might look like *this*:

$$[_{@}\ p, q, \cancel{r} \mid p, s, \cancel{r} \mid r, \cancel{p}, \cancel{q}\ ]$$

As we saw in the presentation of natural deduction proofs for this modal calculus, we could *label* the alternate zones. Let's do that:

$$[_{@}\ p, q, \cancel{r} \mid_{a} p, s, \cancel{r} \mid_{b} r, \cancel{p}, \cancel{q}\ ]$$

So, our syntax adds, to the language of formulas, a category of zone tags. We can then add to our conceptual arsenal the capacity to use these labels in further judgements. It makes sense to say that *in zone b* r *is granted*, or for short 'r : b'. We can think of this as a claim can be made in any zone at all, and introduce it to our vocabulary with this defining rule:

$$\frac{\mathcal{C} \succ A \cdot i}{\mathcal{C} \succ A : i \cdot j} : Df$$

To make the claim that A holds at i, in zone j means no more and no less than to make the claim that A in zone i. ¶ This addition to our vocabulary is conservatively extending and uniquely defining. It has the effect of taking the 'anthropological' perspective of talking *about* the claims we make and the zones in which we make them. We take a 'God's eye view' of the different scenarios we consider, to reflect that talk more explicitly *in* our talk.

» «

Once this addition is made, it is an immediate consequence that these tagged formulas may be included as components of larger formulas. It makes sense to conjoin, disjoin, negate and use first-order quantifiers *on* these formulas, and for these formulas themselves to be tagged. As a result, we are moving toward a fully two-sorted first-order logic, with two basic sorts. Formulas can contain *entity* terms and variables (the original singular terms) and zone tags.

We can also apply modal operators to tagged formulas, but this seems hardly worth the effort, since if the tagged formula $A : i$ is present in one zone, it follows in all others, so all zone-tagged formulas are necessary, if true.

It is worth pausing at this point: adding tagged formulas looks like it comes very close to giving us singular terms referring to *possible worlds*. Such a reading is, of course, *permissible*, but it is in no way *required*. The commitments we have made thus far do not mean that we must treat our tags are referring devices. There is no logical compulsion (other than force of habit) to take these tags to denote possible worlds, where these are understood as some kind of *thing*. Our language already contains a sort (the singular terms) that ranges over things, and there is no requirement that for each tag, we should take it to be referring to any of those things, or, indeed, to anything at all.

In the language under discussion here it is consistent to be an *entity monist* (of necessity, there is only one *thing*) but to take there to be different possibilities. In limit positions modelling such commitments, there is only one entity, but there is more than one *zone*. If we take zone tags to be referring expressions, they cannot refer to *objects*, since there are not enough objects to do the job. This does not settle the issue of whether world-talk should or should not be construed as object-talk, but that the two-type framework in focus here does not force that conclusion upon us.

» «

With the addition of explicit zone tags *in formulas*, it is one small step to allow for *quantification* into zone index position. Let's add zone variables (why not use 'w'?), and with the obvious defining rules for universal and existential quantification into tag position, $\forall w(A : w)$ is equivalent to $\Box A$, and $\exists w(A : w)$ is equivalent to $\Diamond A$. ¶ We are, piece by piece, expanding the expressive capacities of our vocabulary, using only conservatively extending and uniquely defined logical concepts—here, exploiting the labels identifying the different zones for counterfactual supposition—and the result is a richer vocabulary, still inferentially defined, that describes *exactly* the same class of models that we motivated on inferentialist lines. We have moved from a zone-internal vocabulary making claims that differ from one zone to the next, to a zone-neutral *external* vocabulary. This shift incurs no greater ontological commitment than was made previously, but the resulting two-sorted first-order language with object quantification and zone quantification brings us closer to the vocabulary of the two-level type theory Ty_2.

3.3 GOING UP THE LADDER

Our next step at enriching the vocabulary, inferentially, is to add λ abstraction, and to enrich the family of types, beyond having terms for zones (type s), singular terms (type e), sentences (type t), and predicates (type $\langle et\rangle$). The extension to abstraction, and to higher types is easy to motivate inferentially. We already have variables of type e, so let's first discuss abstraction into e position. If our language has the one-place predicates F and G, it is not hard to conceive of an object's being *both* F and G. We can form the sentence $Fx \land Gx$, and this is one step from forming the *complex predicate* $\lambda x(Fx \land Gx)$. Something has this feature if and only if it is both F and G. Predicate abstraction is governed by a straightforward defining rule:

$$\frac{\mathcal{C} \succ A^x_b \cdot i}{\mathcal{C} \succ (\lambda x A)b \cdot i}\ \lambda_{\langle et\rangle}Df$$

To apply a complex predicate to a singular term, substitute the variable bound by the abstraction operator with the term to which it is applied.

What goes for predicate abstraction can go for other types, too. For example, once we have predicates, it is natural to abstract into *predicate* position. Given a predicate variable P (of type $\langle et\rangle$), and a singular term a (type e) the expression $\lambda P\, Pa$ has type $\langle\langle et\rangle t\rangle$. It expects a predicate as an input, and returns a sentence, to the effect that this is a feature borne by the item a. We can generalise: $\lambda P(Pa \lor Pb)$ expresses the higher order property of properties had by *either* a or b. In general, given any type α, we can define the abstraction operator of type $\langle\alpha t\rangle$ generalising the defining rule above:

$$\frac{\mathcal{C} \succ A^P_B \cdot i}{\mathcal{C} \succ (\lambda P\, A)B \cdot i}\ \lambda_{\langle\alpha t\rangle}Df$$

where A has type t, and P and B have type α. ¶ The definition of abstraction at higher types is straightforward on an inferentialist view. Introducing abstraction terms is simply repackaging information that was otherwise available in the prior vocabulary. Indeed, the conservativity and uniqueness of abstraction at all types is straightforward to show, given that formulas with λ-terms can be immediately rewritten without them. ¶ The *power* of abstraction is unleashed, however, when you combine it with *quantification* at all type levels.

Matters are more subtle when we wish to define abstractions of types other than $\langle\alpha t\rangle$ which result in sentences. Gallin defines the general reduction rule for λ terms axiomatising a type-general *identity* relation, where $A \equiv B$ is of type t whenever A and B have the same type. It is more general to impose the defining rule:

$$\frac{\mathcal{C} \succ C\, A^P_B \cdot i}{\mathcal{C} \succ C\,(\lambda P\, A)B \cdot i}\ \lambda_{\langle\alpha\beta\rangle}Df$$

where A has type β, D has type $\langle\beta t\rangle$, and P and B have type α.

Given variables at each type level, it is natural to *generalise* into those variable positions, with exactly the same rules we had previously given, for the existential and universal quantifiers.

$$\frac{\mathcal{C} \succ A \cdot i}{\mathcal{C} \succ \forall P\, A \cdot i}\ \forall_\alpha Df \qquad \frac{\mathcal{C}, A \succ C \cdot i}{\mathcal{C}, \exists P\, A \succ C \cdot i}\ \exists_\alpha Df$$

Here, the variable P has type α, and is required to not occur free in the context $\mathcal{C}$ (or the tag i). ¶ These rules will have their desired effect only in the presence of an underlying principle of SUBSTITUTION, according to which the variable P is inferentially general among terms of type α. With the appropriate substitution guarantees in place, we will be able to make inference steps from $\forall P\, A$ to A^P_B, and from A^P_B to $\exists P\, A$, for any B of type α, as one would hope. ¶ As before, the addition of quantifiers at every type level is uniquely defining, and conservatively extending. However, the conservative extension result comes with significant caveats. ¶ First, the richer structure available when quantifying into higher types (even in the predicate type $\langle et\rangle$) means that quantification is *impredicative*. This means that our usual argument to conservative extension is blocked. The appropriate *left* rule for the universal quantifier has this form:

$$\frac{\mathcal{C}, A^{P}_{B} \succ C \cdot i}{\mathcal{C}, \forall P\, A \succ C \cdot i}\,\forall_{\alpha} L$$

where B is any item of type α. Since this type might involve terms of great logical complexity, it might even contain the formula $\forall P\, A$ as a *subformula*, and hence, there is absolutely no guarantee that the left and right rules only add material and never destroy it, read from top to bottom. (Consider $\forall P(Pa)$, where P has type $\langle et\rangle$. $\lambda x(\forall P(Pa) \wedge Fx)$ also has type $\langle et\rangle$ and so, substituting for the quantifier we get $\lambda x(\forall P(Pa) \wedge Fx)\, a$, which has $\forall P(Pa)$ as a subformula.) ¶ The rules are impredicative, as they stand. It should not be surprising if the higher type quantifier rules behave differently to the first-order quantifiers.

The 'usual argument' goes like this: take the invertible defining rules, and provide equivalent left-right rules which satisfy the subformula property (anything above the inference line is also present below the line). Then show that the only remaining rule in your calculus (the *Cut* rule) may be eliminated, and hence, any derivable sequent can be derived without *Cut*, and so, has a properly *analytic* derivation using only the vocabulary in the endsequent.

Nonetheless, as a family, the λ and quantifier rules at each type *are* conservatively extending over the original two-sorted first-order modal logic. It is not too hard to see why. Any model of the two-sorted first-order modal logic has two domains, a zone domain D_s, and an entity domain D_e (which has a distinguished subset D_e^w for each zone w, of the objects that are taken, in zone w, to exist). For uniformity, we will take there to be a special domain $D_t = \{0, 1\}$ of truth values. A domain $D_{\langle\alpha\beta\rangle}$ for type $\langle\alpha\beta\rangle$ is a set of functions from D_α to D_β. ¶ *Which* functions? The easiest choice, in one sense, is to say *all* such functions. These are the domains that our variables of each type range over, and the standard recursive truth conditions for formulas apply to our enriched language, and they agree, systematically with the truth conditions for our formulas in the original two-sorted first-order language, which was *itself* a conservative extension of the prior modal language without world labels. ¶ Our original simple modal logic, with its modest ontological commitments, sits inside a richer vocabulary, which can be modelled on exactly the same structures. No positions open in the old vocabulary are closed off by appeal to higher-order reasoning. The elaborate edifice of higher types is a metalanguage that the semanticist might use to explicate the compositional features of the languages we wish to understand, and doing so imposes no constraint on any existing first-order modal commitments.

The choice of 'all' such functions is *easy*, but it does involve some serious mathematical and logical baggage. If our entity domain D_e is infinite, the domains $D_{\langle et\rangle}$, $D_{\langle\langle et\rangle t\rangle}$, $D_{\langle\langle\langle et\rangle t\rangle t\rangle}$, . . . climb up the hierarchy of infinite cardinals, and the logic of such 'standard' models is beholden to the commitments of the underlying set theory in which it is formulated. Nothing like this dependence obtains for first-order theories.

Although the 'standard models' of Ty_2 are forbiddingly mathematically rich, and implicated in issues in set theory that seem quite far away from concerns in natural language semantics, these models need play no more than a modelling role, used to assure us that the higher type vocabulary conservatively extends our first-order commitments. There is good evidence that the inferences that actually do the *work*, in formal linguistics, are given in the defining rules for abstraction and the quantifiers [4], so the inferentially defined system suffices.

3.4 WHAT THIS MEANS MEANING MIGHT MEAN

With all this, I have given an inferentialist perspective on a rich and expressive type theory. This language is a structuring tool for the *theorists* to describe and explain compositional patterns that are implicit in how we *use* our vocabulary in our natural languages. There is no requirement that these logical constants at higher types are the kinds of things we find simple or natural as language users. They can be *implicit* in the patterns in our linguistic and conceptual behaviour.

There is no requirement that the everyday language user, competent in the use of the indefinite article 'a', has to understand that is a term of type $\langle\langle et\rangle\langle\langle et\rangle t\rangle\rangle$ or that it means $\lambda P \lambda P' \exists x(Px \wedge P'x)$.

The basic materials in the models for Ty_2 (the ground-level domains of entities, zones and truth values) can be given the same interpretation that the inferentialist gave for models of a first-order modal language. The inhabitants of the entity domain are the possibilia (where commitment to possibilia is explained in terms of commitment to what is possible concerning what exists), and the inhabitants of the *zone* domain simply label the different commitments concerning what is possible. The rest of the edifice built atop these domains is superstructure, classifying this base in various ways. ¶ The *functional* domains of higher types provide a space in which the linguist can describe the capacities and commitments of language users. An item of type $\langle\alpha t\rangle$ corresponds to a *distinction* between items of type α. So, a *predicate* (type $\langle et\rangle$) corresponds to a distinction between entities (between those that have the feature and those that don't). An

I could spend time explaining how appeal to the domain D_t of *truth values* does not incur any special ontological commitment beyond that involved in ruling some things *in* and some things *out*, but everyone seems to have already *got* that point, and no-one is particularly exercised by ontological commitment to D_t.

item of type $\langle\langle et\rangle t\rangle$ corresponds to a distinction between those first-level distinctions, etc. ¶ Given any type α, we have a corresponding *intension* of type $\langle s\alpha\rangle$. Something of this type a choice of something of type α, in each different zone. An item of type $\langle se\rangle$ is a selection of an object, from zone to zone, under counterfactual supposition.

There is much more to be said, but at the very least, we can see how these types might have cognitive and communicative significance. We can *grasp* the meaning of some part of speech (or the corresponding concept), to a lesser or greater extent, as we are able to work with at concept more narrowly or broadly. The particular capacities displayed, of selecting items, categorising them in various ways, categorising those categorisations—and doing so not only in the here-and-now, but also under counterfactual suppositions, as we apply our reasoning and our imagination—seem like just the kinds of cognitive and communicative capacities that are important for creatures like us. ¶ The inferentialist pre-semantics points to how an intensional type theory like Ty_2 provides a useful structuring vocabulary for those capacities, and how it does so without semanticists having to defer to metaphysicians for license to reason in this way. ¶ Truth-conditional semantics is *vindicated* by its inferentialist pre-semantics.

» «

Let me return full circle. In the first lecture, I defined $\to$ like this:

I have made the zone tag explicit, as this rule applies in each zone.

$$\frac{\mathcal{C}, A \succ B \cdot i}{\mathcal{C} \succ A \to B \cdot i}\ \to\!Df$$

When viewed from Ty_2, the material conditional is a lexical item of type $\langle t\langle tt\rangle\rangle$, as it conjoins two sentences. Items of type $\langle t\langle tt\rangle\rangle$ are interpreted in models of Ty_2 as binary truth functions. ¶ Now, the one binary truth function can be presented in many different ways. The lexical entry for material conditional '$\to$' is constrained only by the traditional boolean valuation condition. To grasp a concept with this interpretation, it suffices to affirm $A \to B$ when A is denied or B is affirmed, and to deny $A \to B$ when A is affirmed and B is denied. This is, of course, altogether too strong a constraint, because we might affirm A and deny B and the issue of whether $A \to B$ holds simply does not arise for us. Second, even when the issue does arise, we might find it hard to interpret, due to its complexity. ¶ Competence with a lexical item with an interpretation like this can be acquired in many ways. One is to take $A \to B$ to be a shorthand for $\neg A \vee B$. Another would be to take it as shorthand for $\neg(A \wedge \neg B)$. Yet another would be to give it the defining rule $\to\!Df$ given above. Each will deliver the same pattern of verdicts, when applied correctly, and each would have the same *intension*, as the interpretation is kept fixed under subjunctive suppositions. If a language user has a concept that appears to have the same type, and the pattern of behaviour in the use of that concept robustly differs in extension from this (or it differs in *intension*, as is demonstrated under suitable zone shifts), then this is good evidence that our interlocutor means something other than '$\to$' by this concept.

It can also be lifted to operate on intensions, taking type $\langle\langle st\rangle\langle\langle st\rangle\langle st\rangle\rangle\rangle$.

$v(A \to B) = 0$ iff $v(A) = 1$ and $v(B) = 0$.

QUICK: is $((p \to q) \to p) \to p$ true?

However extension and intension does not exhaust meaning, at least as least as far as inferential behaviour is concerned. While $A \to B$ and $\neg A \vee B$ agree on extension and on intension, they are given different *definitions*. They come to say the same thing by different routes. The rules introducing a concept that is given by definition are *basic* for that concept, and the distinction between basic and derived inferences is important for giving an account of proof, of understanding, and giving an account of how we might acquire concepts. If these issues are important to *semantics*, our semantic theories should address them. Truth-conditional semantic theories, insofar as they represent meaning by extension and intension, focus on the *result* of interpretation, drawing our attention away from the *process*. ¶ Supplementing your truth-conditional semantics with an inferentialist *pre*-semantics means you can keep your representational theory while gaining new insight into its foundations, and a whole host of formal and conceptual tools to apply to address semantic questions.

If I ask my logic students to show that $A \to B$ and $\neg A \vee B$ are equivalent, there is some work that I am asking them to *do*. I am not asking them to show that $A \to B$ and $A \to B$ are equivalent, which is a much simpler task.

WHERE TO, FROM HERE?

Let me end with three gestures to the future.

First, this is just one effort at giving a Belnap-inspired alternative pre-semantic analysis of a familiar semantic system. I would like to see more! Take your favoured foundational theories, and see if they can be understood and analysed in different, non-competing ways. I expect that we will learn things.

Proof theoretical approaches are a natural home for *hyperintensional* distinctions. Even though A and B might be logically equivalent, a proof of A is not necessarily a proof of B. Truth-conditional semantics tends to flatten out distinctions between logically equivalent statements, while they might have different *semantic*, *epistemic*, and *metaphysical* significance. ¶ An inferentialist pre-semantics gives scope for modelling natural and motivated hyperintensional distinctions.

If the proof of the equivalence of A and B is complex and difficult to find, then of course there will be proofs of A that are not also proofs of B.

This entire investigation was motivated by paying attention to the prevailing contours of *logic*, as a foundational discipline with its own distinctive insights. I have sought to make use of both its proof theoretic and model theoretic techniques, because doing so gives us more to work with, as philosophers, and exploring their connections with other foundational issues—in this case, in semantics—gives us new insights which, in turn, means we return to those logical techniques with greater understanding.

Thanks to my Arché colleagues for feedback on an earlier presentation of these lectures, and especially to Franz Berto, Aaron Cotnoir, Justin D'Ambrosio, Viviane Fairbank, Christopher Mastermann, Matteo Nizzardo, Walter Pedriali, Stephen Read and Crispin Wright, for discussions about the ideas herein.

REFERENCES

[1] NUEL BELNAP. "Under Carnap's Lamp: Flat Pre-semantics". *Studia Logica*, 80(1):1–28, 2005.

[2] ELIZABETH COPPOCK AND LUCAS CHAMPOLLION. "Invitation to Formal Semantics". Draft available at `https://eecoppock.info/bootcamp/semantics-boot-camp.pdf`, 2024.

[3] DANIEL GALLIN. *Intensional and Higher-Order Modal Logic: With Applications to Montague Semantics*. North-Holland Pub. Co., Amsterdam; Oxford, 1975.

[4] MICHAEL HAHN AND FRANK RICHTER. "Henkin semantics for reasoning with natural language". *Journal of Language Modelling*, 3(2):513–568, Feb. 2016.

[5] DAVID LEWIS. "General semantics". *Synthese*, 22(1):18–67, 12 1970.

[6] RYAN M. NEFDT. *The Philosophy of Theoretical Linguistics: a Contemporary Outlook*. Cambridge University Press, 2024.

[7] BARBARA H. PARTEE, editor. *Montague Grammar*. Academic Press, 1976.

[8] BARBARA H. PARTEE. "Possible worlds in model-theoretic semantics: a linguistic perspective". In STURE ALLEN, editor, *Possible Worlds in Humanities, Arts, and Sciences: Proceedings of Nobel Symposium 65*, pages 93–123. Walter de Gruyter, Berlin, 1988.

[9] BARBARA H. PARTEE. *Compositionality in Formal Semantics: Selected Papers by Barbara H. Partee*. Explorations in Semantics. Wiley-Blackwell, 2004.

[10] RICHMOND THOMASON, editor. *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press, New Haven, 1974.

[11] TIMOTHY WILLIAMSON. "How did we get here from there? The transformation of analytic philosophy". *Belgrade Philosophical Annual*, 27:7–37, 2014.