# Logic and Philosophical Logic

## Inferentialism and Meaning Underdetermination

A.C. Paseau
alexander.paseau@philosophy.ox.ac.uk
28 January 2019

In the first half of today's class, we looked at Tarski's account of logical constants. His fundamental idea, that the logical constants denote invariant relations, is the basic premise within the *semantic* or *model-theoretic* tradition. The main alternative approach to the logical constants goes by the name of *inferentialism*. In the second half of today's class we'll look at inferentialism and an important problem for it. As with earlier sets of notes for this class, what follows is a springboard for class discussion.

# 1 Inferentialism

Inferentialism is broadly speaking the idea that meaning is use. It is an orientation in the philosophy of language that takes inferential relations to determine the meaning of an expression. In the philosophy of logic, more narrowly, inferentialists take the meaning of the logical constants to be given by the rules characterising them. A typical inferentialist, for example, would maintain that the meaning of the sentential connective 'and' is given by its introduction and elimination rules.[1]

To refresh our memories, let's lay out the usual (classical) introduction and elimination rules for the four propositional connectives $\vee, \neg, \rightarrow$ and $\wedge$. In the case of disjunction, these are:

$$\frac{\Gamma \vdash \phi}{\Gamma \vdash \phi \vee \psi}\ \vee\text{-Intro}_1 \qquad \frac{\Gamma \vdash \phi}{\Gamma \vdash \psi \vee \phi}\ \vee\text{-Intro}_2 \qquad \frac{\Gamma \vdash \phi \vee \psi \qquad \Gamma, \phi \vdash \chi \qquad \Gamma, \psi \vdash \chi}{\Gamma \vdash \chi}\ \vee\text{-Elim}$$

The rules for negation are:

$$\frac{\Gamma, \phi \vdash \psi \qquad \Gamma, \phi \vdash \neg\psi}{\Gamma \vdash \neg\phi}\ \neg\text{-Intro} \qquad \frac{\Gamma, \neg\phi \vdash \psi \qquad \Gamma, \neg\phi \vdash \neg\psi}{\Gamma \vdash \phi}\ \neg\text{-Elim}$$

The conditional rules are:

$$\frac{\Gamma, \phi \vdash \psi}{\Gamma \vdash \phi \rightarrow \psi}\ \rightarrow\text{-Intro} \qquad \frac{\Gamma \vdash \phi \qquad \Gamma \vdash \phi \rightarrow \psi}{\Gamma \vdash \psi}\ \rightarrow\text{-Elim}$$

Finally, here are the conjunction rules:

$$\frac{\Gamma \vdash \phi \qquad \Gamma \vdash \psi}{\Gamma \vdash \phi \wedge \psi}\ \wedge\text{-Intro} \qquad \frac{\Gamma \vdash \phi \wedge \psi}{\Gamma \vdash \phi}\ \wedge\text{-Elim}_1 \qquad \frac{\Gamma \vdash \phi \wedge \psi}{\Gamma \vdash \psi}\ \wedge\text{-Elim}_2$$

[1] Murzi & Steinberger (2017) is an excellent introduction to inferentialism. Some writers prefer the term 'Inferential Role Semantics'.

Suppose now we take propositional logic $\mathsf{PL}$ to consist of a countable infinity of sentence letters $P, Q, R, \cdots$, and the four connectives $\vee, \neg, \rightarrow$ and $\wedge$. The above introduction and elimination rules define a deductive system for $\mathsf{PL}$ that is sound and complete with respect to the usual semantics. Very briefly, this semantics is specified as follows:[2] an assignment is any function from the set of $\mathsf{PL}$'s sentence letters to $\{T, F\}$; a *standard valuation* is any function $v$ from the set of $\mathsf{PL}$-sentences (defined in the usual recursive manner) that extends some assignment and respects the usual rules for the connectives, i.e. $v(\phi \wedge \psi) = T$ iff $v(\phi) = T$ and $v(\psi) = T$, and so on; and semantic consequence is defined as truth-preservation over all standard valuations (i.e. $\Gamma \vDash \delta$ iff for all standard valuations $v$, if $v(\gamma) = T$ for every $\gamma \in \Gamma$ then $v(\delta) = T$).[3] Of course, to put it this way is to adopt the usual 'semantics first' perspective: a good proof system is one that perfectly matches the semantics. As the inferentialist sees it, however, it is the proof system that comes first. Rather than being prior to it, semantics is ultimately determined by the proof system.

## 2 Carnap's categoricity problem

We now consider a prima facie problem for inferentialism. In a nutshell, the problem is that the usual introduction and elimination rules do not determine the correct (classical) semantics for $\mathsf{PL}$. In particular, we cannot recover the usual truth tables for the connectives from our proof system. We'll show that the system does not settle the truth-tables of the four $\mathsf{PL}$ connectives save $\wedge$.

To see this, we need some more terminology. A *generalised valuation* $w$ is *any* function from the set of $\mathsf{PL}$-formulas to the set $\{T, F\}$ of truth-values. NB We do *not* assume that $w$ is a standard valuation, so that its value for complex sentences is determined by its value on atomic sentences; for instance we don't assume that $w(p_0 \vee p_1) = T$ iff $w(p_0) = T$ or $w(p_1) = T$ (or both). We use the symbol $w$, rather than $v$, to mark the difference between generalised valuations and standard ones; standard valuations are thus a proper subset of generalised ones. The point, then, will be to try to recover standard truth clauses from our deductive system. That's why such problems are sometimes said to belong to 'inverse logic'.

We say that $w$ is an *admissible* generalised valuation just when, for any set of $\mathsf{PL}$-sentences $\Gamma$ and $\mathsf{PL}$-sentence $\delta$ such that $\Gamma \vdash \delta$, either $w$ assigns $F$ to at least one member of $\Gamma$ or $w$ assigns $T$ to $\delta$. Alternatively, $w$ is an admissible generalised valuation just when $w$ does not assign $T$ to all members of $\Gamma$ and $F$ to $\delta$ when $\Gamma \vdash \delta$. Finally, say that generalised valuation $w$ *conflicts with the usual truth-table for a connective* if it does not assign the expected truth-value to a compound sentence given the truth-values of its parts; for example $w$ conflicts with the usual truth-table for $\vee$ if, for some $\phi$ and $\psi$, $w(\phi) = F$ and $w(\psi) = F$ but $w(\phi \vee \psi) = T$.

The formal result we now prove is:[4]

> There are admissible generalised valuations $w$ that conflict with the usual truth-tables for $\vee, \neg$ and $\rightarrow$.

[2] Our terminology is a little unusual, for reasons that will become clear.

[3] From now on, we abbreviate 'for every $\gamma \in \Gamma$, $v(\gamma) = T$' by '$v(\Gamma) = T$'.

[4] The result is owed to Carnap (1943); my exposition is based on later presentations.

## 2.1 The argument

We argue for the result by displaying some admissible generalised valuations. Taken together, these show that the deductive system does not pin down the truth-tables for $\vee, \neg$ and $\rightarrow$ uniquely.

Notice first that *any* standard valuation is admissible. To see this, simply invoke the soundness theorem for $\mathsf{PL}$, *viz.* f $\Gamma \vdash \delta$ then $\Gamma \vDash \delta$. Suppose then that $\Gamma \vdash \delta$; by soundness, if $v(\Gamma) = T$ then $v(\delta) = T$, for $v$ any standard valuation.

Consider next the generalised valuation $w_T$ which maps *all* $\mathsf{PL}$-sentences to $T$. 'All' here really does mean *all*: not just all $\mathsf{PL}$-sentence letters but all $\mathsf{PL}$-sentences of any complexity. It is easy to show that this generalised valuation is admissible, whatever the proof system. For the conditional 'if $w_T(\Gamma) = T$ then $w_T(\delta) = T$' holds for any $\Gamma$ and $\delta$, since the antecedent is true for all $\Gamma$ and the consequent is likewise true for all $\delta$. Observe in passing that $w_T$'s admissibility does not turn on specific features of the proof system.

Finally, consider the generalised valuation $w_{Prov}$ which maps all $\mathsf{PL}$-theorems from our earlier deductive system to $T$ and all non-theorems to $F$. A theorem, recall, is anything that is provable from no assumptions; and a non-theorem is anything that's not a theorem. The generalised valuation $w_{Prov}$ is also admissible. For suppose $\Gamma \vdash \delta$ and $w_{Prov}(\Gamma) = T$. Then by definition, all the elements of $\Gamma$ are theorems of our system. Since $\Gamma \vdash \delta$, it follows that $\delta$ is a theorem too. Hence $w_{Prov}(\delta) = T$. So $w_{Prov}$ is an admissible generalised valuation. Observe in passing that $w_{Prov}$'s admissibility, like $w_T$'s, similarly does not hinge on specific features of the proof system.

On this basis, one can show that a standard deductive system does not determine the usual truth-tables for $\vee, \neg$ and $\rightarrow$. Take negation to start with. Since a standard valuation and $w_T$ are both admissible, the first row of negation's truth table is not determined by the system. In a bit more detail: there's an admissible generalised valuation $w_1$ such that $w_1(P) = T$ and $w_1(\neg P) = F$, namely any standard one that assigns $T$ to $P$; and there's an admissible generalised valuation $w_2$ such that $w_2(P) = T$ and $w_2(\neg P) = T$, namely $w_T$. To see that the second row of negation's truth table is not determined by the system either, consider any standard generalised valuation $w_2$ that maps $P$ to $F$ and $w_{Prov}$. Observe that $w_2(P) = F$ and $w_2(\neg P) = T$, whereas $w_{Prov}(P) = F = w_{Prov}(\neg P)$, since neither $P$ nor $\neg P$ is a theorem of the system.

A similar argument shows that the last line of disjunction's truth table is not determined either. For if $w_1$ is a standard valuation and $w_1(P) = w_1(Q) = F$ then $w_1(P \vee Q) = F$; but $w_{Prov}(P) = w_{Prov}(\neg P) = F$, since neither $P$ nor $\neg P$ is a theorem, yet $w_{Prov}(P \vee \neg P) = T$, since $P \vee \neg P$ *is* a theorem.

The same type of argument works for the conditional's fourth line. If $w_1$ is a standard valuation and $w_1(P) = w_1(Q) = F$ then $w_1(P \rightarrow Q) = T$; but $w_{Prov}(P) = w_{Prov}(Q) = F$, since neither $P$ nor $Q$ is a theorem, yet $w_{Prov}(P \rightarrow Q) = F$, since $P \rightarrow Q$ is also not a theorem.

Interestingly, the introduction and elimination rules for conjunction *do* determine conjunction's truth tables. Its introduction rule determines the table's first row and its elimination rules determine the second, third and fourth rows.

## 2.2 Summary

The above considersations create a prima facie problem for inferentialists. For they seem to show that inference rules do not determine the meaning of the logical constants, at least not the classical ones.

Inferentialists may be able to argue that various non-standard yet admissible valuations should be ruled out. But making that case for $w_{Prov}$ and $w_T$ seems difficult. It's hard to see what is wrong, inferentially speaking, with a subject who accepts all propositions, and with a subject who accepts all and only theorems.

To spell out the latter claim, consider a subject $S$ who accepts all and only theorems of our interpreted $\mathsf{PL}$ system and reasons hypothetically using its rules. For example, she accepts $P \to P$ (a theorem of her system) but not $P$ (a non-theorem), and reasons hypothetically from acceptance of $P$ to acceptance of $P \vee Q$ (seeing as $P \vee Q$ follows from $P$ in her deductive system). $S$'s inferential dispositions are captured by $w_{Prov}$, which assigns $T$ to all propositional theorems and $F$ to all non-theorems, which as saw does not respect the classical truth-tables for disjunction, negation and the conditional. Our imagined subject $S$, who is disposed to accept $\phi$ iff $w_{Prov}(\phi) = T$ iff $\vdash \phi$, is epistemically extremely conservative: she accepts only propositional theorems and does not accept any non-theorems.[5] But there does not seem to be anything logically or inferentially amiss with her. Her inferential practice—her dispositions to accept certain sentences given others[6]—seems conceivable. If so, inferential practice underdetermines the propositional connectives' meaning.

In closing this section, we note that the categoricity problem is likely to persist when we move to stronger classical logics than $\mathsf{PL}$. For the analogues of the valuations $w_{Prov}$ and $w_T$ can also be given there. They will be admissible in these contexts too, and promise to create analogous problems for the recovery of the correct clauses for these logics' connectives.

In class, we'll consider some responses by the inferentialist. These include: reject classical semantics; take a 'bilateral' approach; adopt multiple-conclusion logic. I'll briefly introduce each in class before throwing open the discussion. To discuss the last two in an informed way, you need to know what bilateral rules and multiple-conclusion rules look like, so I append a version of each below in case you haven't seen them before.

# 3 Bilateral Rules

A key reference for bilateralism is Smiley (1996). The rules and explanation below are taken from Rumfitt (2002, pp. 800-2).

When $A$ is a declarative sentence, the *signed sentence* $+A$ abbreviates the question-answer pair 'Is it the case that $A$? Yes', whereas $-A$ abbreviates the question-answer pair 'Is it the case that $A$? No'. The positively signed formula $+A$ is correct if $A$ is true and incorrect if $A$ is not true; the negatively-signed formula $-A$ is correct if $A$ is not true and incorrect if $A$ is true. NB It is important not to confuse the sentential operator $\neg$, which can be iterated indefinitely, with the force

[5] 'Under an interpretation' understood throughout.

[6] Including as a special case her dispositions to accept certain sentences given no others.

operator $-$, which cannot be iterated. A set of correctness-preserving bilateralist rules may then be given as follows:

$$\frac{+A \qquad +B}{+(A \wedge B)} + \wedge I \qquad \frac{+(A \wedge B)}{+A} + \wedge E1 \qquad \frac{+(A \wedge B)}{+B} + \wedge E2$$

$$\frac{-A}{-(A \wedge B)} - \wedge I1 \qquad \frac{-B}{-(A \wedge B)} - \wedge I2$$

$$\frac{-(A \wedge B) \qquad \begin{matrix}[-A] \\ \vdots \\ \phi\end{matrix} \qquad \begin{matrix}[-B] \\ \vdots \\ \phi\end{matrix}}{\phi} - \wedge E$$

$$\frac{+A}{+(A \vee B)} + \vee I1 \qquad \frac{+B}{+(A \vee B)} + \vee I2$$

$$\frac{+(A \vee B) \qquad \begin{matrix}[+A] \\ \vdots \\ \phi\end{matrix} \qquad \begin{matrix}[+B] \\ \vdots \\ \phi\end{matrix}}{\phi} + \vee E$$

$$\frac{-A \qquad -B}{-(A \vee B)} - \vee I \qquad \frac{-(A \vee B)}{-A} - \vee E1 \qquad \frac{-(A \vee B)}{-B} - \vee E2$$

$$\frac{\begin{matrix}[+A] \\ \vdots \\ +B\end{matrix}}{+(A \to B)} + \to I \qquad \frac{+(A \to B) \qquad +A}{+B} + \to E$$

$$\frac{+A \qquad -B}{-(A \to B)} - \to I \qquad \frac{-(A \to B)}{+A} - \to E1 \qquad \frac{-(A \to B)}{-B} - \to E2$$

$$\frac{-A}{+(\neg A)} +\neg I \qquad \frac{+(\neg A)}{-A} +\neg E$$

$$\frac{+A}{-(\neg A)} -\neg I \qquad \frac{-(\neg A)}{+A} -\neg E$$

A relatively straightforward argument shows that any valuation compatible with the positive and negative rules for constant $c$ must respect the usual truth-table for $c$.

# 4 Multiple-Conclusion Logic

A key reference for the multiple-conclusion approach to the logical constants is Hacking (1979). The rules below are taken from Garson (2013, p. 164).

We say that a generalised valuation (defined as any map from the set of PL-formulas to the set of truth-values, as above) satisfies a multiple-conclusion sequent $\Gamma \vdash \Delta$ just when it either assigns $F$ to a member of $\Gamma$ or it assigns $T$ to a member of $\Delta$. (Alternatively, just when it does not assign $T$ to all members of $\Gamma$ and $F$ to all members of $\Delta$.) A generalised valuation $v$ is compatible with some multiple-conclusion sequent rules on condition that it satisfy the sequent(s) below the line if it satisfies the sequent(s) above the line.

Multiple-conclusion sequents usually contain two sorts of rules: structural rules; and a pair of left and right rules for each logical constant, the analogues of operational rules. The structural rules are usually taken to be:

Hypothesis: $\Gamma \vdash \Delta$, whenever some formula is in both $\Gamma$ and $\Delta$

Left Dilution: $\dfrac{\Gamma \vdash \Delta}{\Gamma, \Gamma^* \vdash \Delta}$

Right Dilution: $\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, \Delta^*}$

Cut: $\dfrac{\Gamma \vdash \phi, \Delta \qquad \Gamma^*, \phi \vdash \Delta^*}{\Gamma, \Gamma^* \vdash \Delta, \Delta^*}$

The left and right rules for the four constants mentioned are:

$$\dfrac{\Gamma, \phi \vdash \Delta \qquad \Gamma, \psi \vdash \Delta}{\Gamma, \phi \vee \psi \vdash \Delta}\ \vee\text{-Left} \qquad \dfrac{\Gamma \vdash \phi, \psi, \Delta}{\Gamma \vdash \phi \vee \psi, \Delta}\ \vee\text{-Right}$$

$$\dfrac{\Gamma \vdash \phi, \Delta}{\Gamma, \neg\phi \vdash \Delta}\ \neg\text{-Left} \qquad \dfrac{\Gamma, \phi \vdash \Delta}{\Gamma \vdash \neg\phi, \Delta}\ \neg\text{-Right}$$

$$\dfrac{\Gamma \vdash \phi, \Delta \qquad \Gamma, \psi \vdash \Delta}{\Gamma, \phi \rightarrow \psi \vdash \Delta}\ \rightarrow\text{-Left} \qquad \dfrac{\Gamma, \phi \vdash \psi, \Delta}{\Gamma \vdash \phi \rightarrow \psi, \Delta}\ \rightarrow\text{-Right}$$

$$\dfrac{\Gamma, \phi, \psi \vdash \Delta}{\Gamma, \phi \wedge \psi \vdash \Delta}\ \wedge\text{-Left} \qquad \dfrac{\Gamma \vdash \phi, \Delta \qquad \Gamma \vdash \psi, \Delta}{\Gamma \vdash \phi \wedge \psi, \Delta}\ \wedge\text{-Right}$$

A relatively straightforward argument shows that any valuation compatible with the right and left rules for constant $c$ must respect the usual truth-table for $c$.

# References


[1] R. Carnap (1943), *Formalization of Logic*, Harvard University Press.

[2] J. Garson (2013), *What Logics Mean*, Cambridge University Press.

[3] I. Hacking (1979), 'What is logic?', *The Journal of Philosophy* 76, pp. 285-319.

[4] I. Rumfitt (2000), '"Yes" and "No"', *Mind* 109, pp. 781-823.

[5] T.J. Smiley (1996), 'Rejection', *Analysis* 56, pp. 1-9.

[6] F. Steinberger & J. Murzi (2017), 'Inferentialism', *Blackwell Companion to Philosophy of Language*, Wiley Blackwell, pp. 197-224.