

Logic Informed

By

Justin Bledin

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Philosophy

in

Logic and the Methodology of Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John MacFarlane, Chair

Professor Niko Kolodny

Professor Seth Yalcin

Professor Paolo Mancosu

Professor Hans Sluga

Fall 2013

Abstract

Logic Informed

by

Justin Bledin

Doctor of Philosophy in Logic and the Methodology of Science

University of California, Berkeley

Professor John MacFarlane, Chair

I develop an informational conception of logic as a science fundamentally concerned not with truth but with information. Facts about logical validity, on this conception, tell us about the structure of the bodies of information that we generate, encounter, absorb, and exchange as we interact with one another and learn about our world. I also investigate the normative role of logic in our epistemic practices. In particular, I argue against the widespread idea that there are rational requirements to have logically coherent beliefs that are not merely epiphenomenal on evidential norms.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction: | |
| | Rival Conceptions of Logic | 1 |
| 2 | The Truth Preservation View of Logic | 6 |
| 3 | Against Truth Preservation | 13 |
| 3.1 | \models_{Tr} Versus Good Argument | 13 |
| 3.2 | Sophisticated Truth Preservation | 15 |
| 3.3 | Truth Preservation Versus Good Argument | 19 |
| 3.3.1 | Trouble With Descriptivism | 20 |
| 3.3.2 | Expressivism | 22 |
| 3.3.3 | Modality With Content | 24 |
| 4 | The Informational View of Logic | 29 |
| 4.1 | Responses | 29 |
| 4.2 | A Better Option | 32 |
| 5 | <i>Modus Ponens</i> Defended | 35 |
| 5.1 | Kolodny and MacFarlane | 35 |
| 5.1.1 | <i>Reductio</i> | 35 |
| 5.1.2 | Constructive Dilemma | 39 |
| 5.2 | Willer | 42 |
| 5.2.1 | Attack | 43 |
| 5.2.2 | Defense | 44 |
| 6 | The Myth of Logical Coherence | 49 |
| 6.1 | Bridging Logic and Epistemology | 50 |
| 6.2 | Problems for Logical Requirements | 53 |
| 6.3 | Kolodny’s Error Theory for Non-Contradiction | 57 |
| 6.3.1 | Explaining Violation _N | 57 |
| 6.3.2 | Explaining Satisfaction _N | 58 |
| 6.4 | Kolodny’s Error Theory for Single-Premise Closure | 59 |

| | | |
|----------|---|-----------|
| 6.4.1 | Explaining Violation _{SPC} | 59 |
| 6.4.2 | Explaining Satisfaction _{SPC} | 60 |
| 6.5 | Pressure Points | 61 |
| 6.6 | Towards Another Error Theory | 65 |
| 6.7 | A Simple Error Theory for Non-Contradiction | 66 |
| 6.8 | A Simple Error Theory for Single-Premise Closure | 67 |
| 6.9 | Virtues | 68 |
| 6.10 | Conclusion | 70 |
| 7 | Conclusion | 71 |
| A | Two Kinds of Supposition | 75 |
| A.1 | Lossy Supposition | 75 |
| A.2 | Lossless Supposition | 78 |
| A.3 | Conditional Declaratives | 80 |
| A.4 | Conditional Interrogatives | 85 |
| B | The Natural Deduction System Info | 89 |
| B.1 | Languages | 89 |
| B.2 | Rules of Info | 91 |
| B.3 | Soundness | 93 |
| B.4 | Completeness | 95 |

Chapter 1

Introduction: Rival Conceptions of Logic

What is logic? Two answers by Gottlob Frege dominate contemporary discussion.

First: Logic is a descriptive science, a body of truths about *truth*. Frege opens his essay *Thoughts* [1918] with this characteristic passage:¹

Just as ‘beautiful’ points the way for aesthetics and ‘good’ for ethics, so do words like ‘true’ for logic. All sciences have truth as their goal; but logic is also concerned with it in a quite different way: logic has much the same relation to truth as physics has to weight or heat. To discover truths is the task of all sciences; it falls to logic to discern the laws of truth...Here of course it is not a matter of what happens but of what is. (p. 351)

In keeping with this characterization, most logicians and philosophers today think that a deductive argument is logically valid just in case it necessarily preserves truth by virtue of its logical form.²

Second: Logic is a normative science, a body of rules that govern our thinking. Frege offers this other characterization in his *Grundgesetze der Arithmetik* [1893]:³

¹See also Frege [1879-1891], p. 3, and [1897], p. 128.

²The rider ‘by virtue of logical form’ is meant to preclude arguments like this from counting as logically valid:

(P1) Alfred is a bachelor.

(C) Alfred is an unmarried man.

But drawing a principled divide between the logical and non-logical expressions of a language is a difficult open problem in the philosophy of logic. See MacFarlane [2009] for a good critical survey of various approaches to this demarcation problem.

³See also Frege [1893], pp. 14, 15, and [1897], pp. 128, 146, 149.

It will be granted by all at the onset that the laws of logic ought to be guiding principles for thought in the attainment of truth, yet this is only too easily forgotten, and here what is fatal is the double meaning of the word ‘law.’ In one sense a law asserts what is; in the other it prescribes what ought to be. Only in the latter sense can the laws of logic be called ‘laws of thought’: so far as they stipulate the way in which one ought to think. Any law asserting what is, can be conceived as prescribing that one ought to think in conformity with it, and is thus in that sense a law of thought. This holds for laws of geometry and physics no less than for laws of logic. The latter have a special title to the name ‘laws of thought’ only if we mean to assert that they are the most general laws, which prescribe universally the way in which one ought to think if one is to think at all. (p. 12)

Whereas Euclid’s principles are in one sense norms for thinking about the geometry of space, and similarly the laws of physics are norms for thinking about the physical world, the laws of logic are constitutive norms for thought *as such*.⁴ That is, something not subject to the laws of logic just isn’t thinking. Though some philosophers will surely balk at this constitutivity claim, most still think that logic is normative for thought in some form or another.⁵

For Frege, of course, these two conceptions of logic are compatible. The above quote from *Thoughts* continues,

From the laws of truth there follow prescriptions about asserting, thinking, judging, inferring. (p. 351)

Thus, on what I think is the most natural reading of Frege, logical laws have descriptive content about truth. They say nothing about how we

⁴cf. MacFarlane [2002].

⁵For example, Barwise and Etchemendy [1999] write on the first page of their popular logic textbook:

All rational inquiry depends on logic, on the ability of people to reason correctly most of the time, and, when they fail to reason correctly, on the ability of others to point out the gaps in their reasoning.

They continue a few pages later:

Rational inquiry, in our sense, is not limited to academic disciplines, and so neither are the principles of logic. If your beliefs about a close friend logically imply that he would never spread rumors behind your back, but you find out that he has, then your beliefs need revision. Logical consequence is central, not only to the sciences, but to virtually every aspect of everyday life.

ought to reason or what we ought to believe. However, the laws of logic still imply prescriptions for thought and related intentional activity. In this regard too, Frege’s view remains popular. In fact, something like his picture of logic has a good claim to being the ‘standard view’ on the contemporary philosophical scene. Logically valid arguments, it is said, necessarily preserve truth. But fans of truth preservation needn’t deny that logic has normative import, nor that the normative component of logic can play an important role in demarcating it from closely related disciplines like geometry and psychology. They must only recognize that logical laws are not themselves norms for thought. Bridge principles are required at the logic-epistemology interface.⁶

To be careful, we should really distinguish between three prevailing conceptions of logical validity. Where $\varphi_1, \dots, \varphi_n, \psi$ are sentences in a formal or informal language,⁷

Valid_T The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if it is impossible for each of the premises $\varphi_1, \dots, \varphi_n$ to be true and for the conclusion ψ to be false by virtue of their logical form.⁸

Valid_G The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if it is a *good* deductive argument that we do well to make in both categorical and hypothetical deliberative contexts by virtue of its logical form.

Valid_N The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if it plays a special normative role in our epistemic and/or linguistic practices—for instance, the argument is valid just in case, subject to meeting certain qualifications, it determines a normative constraint of deductive cogency on thought *as such*.

⁶That such bridge principles are required is one of the key lessons of Gilbert Harman’s *Change in View* [1986]. Here are a couple candidate synchronic principles that bind at each moment in time:

Consistency: Where $\varphi_1, \dots, \varphi_n$ are logically inconsistent, rationality requires you either not to believe that φ_1 is true, ..., or not to believe that φ_n is true.

Closure: Where $\varphi_1, \dots, \varphi_n$ logically imply ψ , rationality requires you either not to believe that φ_1 is true, ..., not to believe that φ_n is true, or to believe that ψ is true. Such logical coherence requirements are the stars of Chapter 6.

⁷Many philosophers think that logical validity applies primarily to series of propositions—the content expressed by sentences in context. But I’ll work with sentences for reasons I discuss in Chapter 3.

⁸Likewise, φ is a logical truth iff it is impossible for φ to be false by virtue of its logical form, $\varphi_1, \dots, \varphi_n$ are logically consistent iff it is possible for $\varphi_1, \dots, \varphi_n$ to be jointly true by virtue of their logical form, and so on.

These conceptions are not generally regarded as serious rivals; while Valid_T is often taken as basic, many philosophers think that all three conceptions of validity extensionally coincide. From the third-person standpoint of appraisal, the arguments which necessarily preserve truth are precisely those which we should positively assess as good deductive arguments. From the first-person standpoint of deliberation and the related second-person standpoint of advice, these good arguments are precisely those which inform, in this or that special sense, what we ought to believe.⁹

But can we hold onto this harmonious standard view? Is logic a descriptive science of truth that tells us which deductive arguments are good and that gives rise to norms for thought? After clarifying what I take to be the standard truth preservation view in Chapter 2, I consider various inferences in Chapter 3 that militate against it. Given what one of our best formal semantic theories says about the semantic values of informational modal operators and the indicative conditional, some good deductive arguments involving these *informational constants* are arguably not truth preserving—that is, Valid_G conflicts with something even weaker than Valid_T , *viz.*, the idea that *material* truth preservation is a necessary condition for validity. Even worse, sentences involving these informational constants are arguably not, properly speaking, true or false. So the truth preservation view seems thoroughly ill-suited to explain the goodness of these deductive arguments. We seem to face a difficult choice: we can either maintain that these good deductive arguments are logically valid, and abandon the view that logically valid arguments necessarily preserve truth; or we can maintain Valid_T but reject Valid_G .

In fact, I argue that we can take the first fork with few tears. I propose in Chapter 4 that logic is not the science of what forms of inference necessarily preserve truth by virtue of logical form, but rather a descriptive science whose proper subject is *information*. Facts about logical validity, on this informational view, tell us about the structure of the bodies of information that we generate, encounter, absorb, and exchange as we interact with one another and learn about our world. Roughly put,

⁹The evaluative and normative dimensions of logic are often conflated. But I think *good deductive argument* is best understood as an evaluative notion applicable in the third-person standpoint, not a normative notion applicable in the first-person and second-person standpoints. It is one thing to know which inferences we do well to make in our reasoning. It is another thing to know what we ought to believe at a particular time or how we ought to revise our beliefs over time. More on this in Chapter 6.

Valid_I The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if information with the structural features (to be made precise in due course) corresponding to each of the premises $\varphi_1, \dots, \varphi_n$ also has the structural feature corresponding to conclusion ψ by virtue of the logical form of these sentences.

This informational conception of validity lines up with the standard conception defined in terms of truth preservation over simple languages without informational modals and the indicative conditional. But logical validity, understood informationally, also dovetails with good deductive argument over rich languages with these informational constants—or so I'll argue. In Chapter 5, I consider episodes of deductive argumentation that purportedly show that *modus ponens* for the indicative—a valid form of argument on the informational view of logic—is unreliable in hypothetical contexts. I argue that these episodes show no such thing; they rather show, *inter alia*, that methods of deductive argumentation involving hypothetical reasoning like *reductio ad absurdum* do not always fit the classical model when performed in a language with informational modals and the indicative conditional.

In Chapter 6, I turn to the normative conception of logic Valid_N. If Valid_I and Valid_G extensionally coincide, we can still ask: inside the first-person deliberative perspective, in what special sense is a logically valid inference like *modus ponens* for the indicative deontic or response guiding? Here my conclusion is discouraging; I argue that the normative role of logic in our epistemic practices is not as strong as is commonly thought. Rational requirements to have beliefs that are consistent and closed under logical consequence, whatever else might be the case, have been regarded by some philosophers as secure pillars of normativity. But I follow Kolodny [2007] and throw some cold water on this idea. Since logical coherence requirements are mysterious within the deliberative standpoint, I present a theory of error for their normativity. Although we needn't have logically consistent and closed beliefs *per se*, I argue that by attending to the informational background of theoretical deliberation, we can nonetheless explain the rational pressure we feel to be coherent.

Chapter 2

The Truth Preservation View of Logic

I use the definite description ‘the truth preservation view’ to denote a cluster of widespread intuitions about the *informal* concept of logical validity.¹ The most basic intuition, of course, is that a logically valid argument never has true premises and a false conclusion. But two further intuitions sharpen this core condition. The first is that validity involves a modal element: it is *impossible* for each of the premises of a logically valid argument to be true and for the conclusion to be false. The second is that a logically valid argument preserves truth *by virtue of the logical form* of the sentences in the argument, and not due to the meaning of any non-logical symbols. These intuitions all come together in Valid_T from Chapter 1: the argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if it is impossible for each of $\varphi_1, \dots, \varphi_n$ to be true and for ψ to be false by virtue of their logical form.²

¹Tarski opens his seminal essay on logical consequence [1936a] with this remark on the subject of mathematical logic:

The concept of logical consequence is one whose introduction into the field of strict formal investigation was not a matter of arbitrary decision on the part of this or that investigator; in defining this concept, efforts were made to adhere to the common usage of the language of everyday life. (p. 409)

Let me be clear that what I have in mind here is not the “common usage” of people on the street (I’m doubtful that this “common usage” even exists), but the informal concept of logical validity that has been influential in philosophy and mathematics since Aristotle.

²Even among logicians and philosophers who accept this definition, there is room for disagreement. Is the modality alethic, metaphysical, or epistemic? Are higher order quantifiers, say, part of logical form? ‘The truth preservation view’ is best regarded as an umbrella term covering the many possible ways to make this definition precise.

Relevance logicians like Anderson, Belnap, and Dunn [1992] also insist that the premises of a logically valid argument must be relevant to its conclusion. However, they argue that this requirement of relevance is not a separate virtue but is actually required to ensure a kind of truth preservation. See Lewis [1982] for good discussion.

In mathematical logic, this informal characterization of validity is typically analyzed in terms of *truth in a model*.³ Open up just about any logic textbook and you'll see something like this:

Valid_M The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if there is no model \mathcal{M} for formal language \mathcal{L} such that the translations of $\varphi_1, \dots, \varphi_n$ into \mathcal{L} are all true in \mathcal{M} but the translation of ψ into \mathcal{L} is false in \mathcal{M} .

where \mathcal{L} is a formal language that purportedly makes the logical form of $\varphi_1, \dots, \varphi_n, \psi$ explicit, and a model \mathcal{M} for \mathcal{L} is, roughly, something that provides enough information to determine the extensions of all well formed sentences $S_{\mathcal{L}}$ of this language. In sentential logic, a model is (basically) just a reference row of a truth table. In classical first-order logic, a model is a non-empty domain of individuals and an interpretation function that maps constants in \mathcal{L} to individuals in the domain and maps predicates in \mathcal{L} to sets of individuals. In intuitionistic logic, a model is a Kripke tree with valuations at each node. And so forth. By varying our formal languages and models, we have generated a large family of formal characterizations of logical validity.

But while logicians and philosophers agree on which arguments count as logically valid on this or that formal characterization, they disagree on which of the formal notions explicate the *genuine* informal notion of logical validity. That is, they disagree on which of the formal notions extensionally coincide with our pre-theoretic, intuitive notion of logical validity. Having an extensionally adequate formal explication of validity would certainly be useful. By investigating it, we could learn things about our target, the informal notion.

Let us first assume that \mathcal{L} is the language of sentential logic with the following symbols:⁴ sentence letters A, B, C, \dots , the contradiction symbol \perp , the logical connectives $\neg, \vee, \wedge, \supset, \equiv$, and parentheses $()$.

A model $\mathcal{M} : At_{\mathcal{L}} \mapsto \{T, F\}$ for \mathcal{L} is an assignment of truth values to all sentence letters $At_{\mathcal{L}}$ in this language. Since the logical connectives are truth functional, \mathcal{M} is easily extended to the full interpretation function $\llbracket \cdot \rrbracket_{\mathcal{M}} : S_{\mathcal{L}} \mapsto \{T, F\}$ for \mathcal{L} mapping all sentences $S_{\mathcal{L}}$ in this language to truth values:

³Besides this model-theoretic semantic approach, logicians also study syntactic characterizations of validity in proof systems.

⁴Going forward, I'll be loose about use and mention, usually omitting Quinean quasi-quotes and the like.

$$\begin{aligned}
\llbracket A \rrbracket_{\mathcal{M}} = T & \text{ iff } \mathcal{M}(A) = T \\
\llbracket \perp \rrbracket_{\mathcal{M}} = T & \text{ iff } 0 = 1 \\
\llbracket \neg\varphi \rrbracket_{\mathcal{M}} = T & \text{ iff } \llbracket \varphi \rrbracket_{\mathcal{M}} = F \\
\llbracket \varphi \vee \psi \rrbracket_{\mathcal{M}} = T & \text{ iff } \llbracket \varphi \rrbracket_{\mathcal{M}} = T \text{ or } \llbracket \psi \rrbracket_{\mathcal{M}} = T \\
\llbracket \varphi \wedge \psi \rrbracket_{\mathcal{M}} = T & \text{ iff } \llbracket \varphi \rrbracket_{\mathcal{M}} = T \text{ and } \llbracket \psi \rrbracket_{\mathcal{M}} = T \\
\llbracket \varphi \supset \psi \rrbracket_{\mathcal{M}} = T & \text{ iff } \llbracket \varphi \rrbracket_{\mathcal{M}} = F \text{ or } \llbracket \psi \rrbracket_{\mathcal{M}} = T \\
\llbracket \varphi \equiv \psi \rrbracket_{\mathcal{M}} = T & \text{ iff } \llbracket \varphi \rrbracket_{\mathcal{M}} = \llbracket \psi \rrbracket_{\mathcal{M}}
\end{aligned}$$

Sentence φ is true in \mathcal{M} if and only if $\llbracket \varphi \rrbracket_{\mathcal{M}} = T$. Logical validity is defined in terms of truth in \mathcal{M} :

Def 1. The argument from $\varphi_1, \dots, \varphi_n$ to ψ is valid, $\{\varphi_1, \dots, \varphi_n\} \models \psi$, just in case there is no model \mathcal{M} where $\llbracket \varphi_1 \rrbracket_{\mathcal{M}} = \dots = \llbracket \varphi_n \rrbracket_{\mathcal{M}} = T$ and $\llbracket \psi \rrbracket_{\mathcal{M}} = F$.⁵

For example, this argument is logically valid:

(P1) Bo Peep lost her sheep.

(C) Either Bo Peep lost her sheep or she lost her dogs.

But this argument is not:

(P1) Either Jack went up the hill or he went to the market.

(C) Jack went to the market.

If $\mathcal{M}(A) = T$, then $\llbracket A \vee B \rrbracket_{\mathcal{M}} = T$. However, if $\llbracket A \vee B \rrbracket_{\mathcal{M}} = T$, then it might still be that $\mathcal{M}(B) = F$.

Of course, there is surely more to logical form than *sentential form*. In sentential logic, only expressions like ‘It is not the case that...’ and ‘Either...or...’ that get translated using \neg , \vee , \wedge , \supset , and \equiv count as the logical part of our language. However, the catalog of logical constants presumably also includes at least first-order quantification and some intensional operators.⁶ Arguments that necessarily preserve truth by virtue of their sentential form are logically valid in the genuine informal sense. But many genuinely valid arguments are sententially invalid. To have any hope of explicating the target informal notion of logical validity, we must consider more sophisticated formal languages than the language of sentential logic.

As a step in this direction, let us now assume that in addition to sentence letters, \perp , the sentential connectives, and parentheses, \mathcal{L} also

⁵For a cleaner exposition, I’ll often use $\varphi_1, \dots, \varphi_n, \psi$ to designate both sentences in English and their translations into a formal language.

⁶Again, the catalog of logical constants is controversial. But I assume in what follows, along with many logicians and philosophers, that informational modals and indicative conditionals make the list.

includes *informational* necessity \Box and possibility \Diamond operators, and the indicative conditional \Rightarrow .⁷

A model $\mathcal{M} = \langle \mathcal{W}, \mathcal{V} \rangle$ for \mathcal{L} now consists of a set of possible worlds \mathcal{W} and a function $\mathcal{V} : At_{\mathcal{L}} \times \mathcal{W} \mapsto \{T, F\}$ mapping each sentence letter $\varphi \in At_{\mathcal{L}}$ and world $w \in \mathcal{W}$ to a truth value. To give a semantics for the full language, I follow Yalcin [2007] and [2011] and Kolodny and MacFarlane [2010] and evaluate sentences in $S_{\mathcal{L}}$ for truth relative both to a world $w \in \mathcal{W}$ and to an *information state* $i \in 2^{\mathcal{W}}$. A recursive specification of truth at *index* $\langle w, i \rangle$ lifts \mathcal{V} to the full interpretation function $\llbracket \cdot \rrbracket_{\mathcal{M}} : S_{\mathcal{L}} \times \mathcal{W} \times 2^{\mathcal{W}} \mapsto \{T, F\}$ for \mathcal{L} mapping $\varphi \in S_{\mathcal{L}}$, $w \in \mathcal{W}$, and $i \in 2^{\mathcal{W}}$ to a truth value. $\llbracket \cdot \rrbracket_{\mathcal{M}}^{w,i}$ is obtained by holding w and i fixed.⁸

The clauses for sentence letters, \perp , and the sentential connectives are like before:

$$\begin{aligned} \llbracket A \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } \mathcal{V}(A, w) = T \\ \llbracket \perp \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } 0 = 1 \\ \llbracket \neg\varphi \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } \llbracket \varphi \rrbracket_{\mathcal{M}}^{w,i} = F \\ \llbracket \varphi \vee \psi \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } \llbracket \varphi \rrbracket_{\mathcal{M}}^{w,i} = T \text{ or } \llbracket \psi \rrbracket_{\mathcal{M}}^{w,i} = T \\ \llbracket \varphi \wedge \psi \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } \llbracket \varphi \rrbracket_{\mathcal{M}}^{w,i} = T \text{ and } \llbracket \psi \rrbracket_{\mathcal{M}}^{w,i} = T \\ \llbracket \varphi \supset \psi \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } \llbracket \varphi \rrbracket_{\mathcal{M}}^{w,i} = F \text{ or } \llbracket \psi \rrbracket_{\mathcal{M}}^{w,i} = T \\ \llbracket \varphi \equiv \psi \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } \llbracket \varphi \rrbracket_{\mathcal{M}}^{w,i} = \llbracket \psi \rrbracket_{\mathcal{M}}^{w,i} \end{aligned}$$

However, the semantic clauses for the informational modal operators and indicative conditional are more interesting since the information state parameter i in the index comes into action:

$$\begin{aligned} \llbracket \Box\varphi \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } \forall w' \in i (\llbracket \varphi \rrbracket_{\mathcal{M}}^{w',i} = T) \\ \llbracket \Diamond\varphi \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } \exists w' \in i (\llbracket \varphi \rrbracket_{\mathcal{M}}^{w',i} = T) \end{aligned}$$
⁹

⁷As I'll discuss in a moment, sentences involving informational modal operators are evaluated relative to an information state (a set of possible worlds). These modals are often called *epistemic* modals since the relevant information state often explicates a knowledge or belief state.

⁸Lewis [1980], drawing on Kaplan [1989], argues that a sentence φ in a natural language must be evaluated relative both to the *context* c in which φ is used, and to an *index*, an n -tuple of features of context. But given the formal language \mathcal{L} under consideration, context dependence and much index dependence can be ignored; the index $\langle w, i \rangle$ will do.

⁹The main difference between our compositional semantics and the standard textbook semantics for \Box and \Diamond is effectively that the information state parameter i in the index $\langle w, i \rangle$ is not a function of the world parameter w . On the textbook semantics, a model $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$ for a modal language like \mathcal{L} also includes a binary

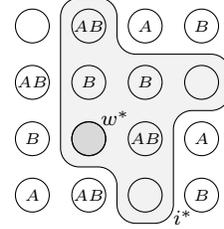


Fig 1. An index $\langle w^*, i^* \rangle$

$$\llbracket \varphi \Rightarrow \psi \rrbracket_{\mathcal{M}}^{w,i} = T \quad \text{iff} \quad \forall w' \in i + \varphi (\llbracket \psi \rrbracket_{\mathcal{M}}^{w',i+\varphi} = T)$$

where $i + \varphi$ appearing in the clause for the indicative is the largest subset $i' \subseteq i$ such that $\forall w \in i' (\llbracket \varphi \rrbracket_{\mathcal{M}}^{w,i'} = T)$.¹⁰ Intuitively: ‘Colonel Mustard must have done it’ is true at index $\langle w, i \rangle$ just in case all of the worlds in i are worlds in which Colonel Mustard did it. ‘Colonel Mustard might have done it’ is true at index $\langle w, i \rangle$ just in case Colonel Mustard did it at some of these worlds. And ‘If Colonel Mustard did it then he used the candlestick’ is evaluated for truth at index $\langle w, i \rangle$ by first ‘adding’ the information that Colonel Mustard did it to the existing stock of information i and then ‘checking’ whether Colonel Mustard used the candlestick at all worlds in this updated information state. Given the index $\langle w^*, i^* \rangle$ in Fig 1, for example, $\llbracket \Box A \rrbracket_{\mathcal{M}}^{w^*,i^*} = F$, $\llbracket \Diamond A \rrbracket_{\mathcal{M}}^{w^*,i^*} = T$, and $\llbracket A \Rightarrow B \rrbracket_{\mathcal{M}}^{w^*,i^*} = T$.

What about the formal logical consequence relation? This is where things get more interesting—there are multiple *prima facie* attractive options for \models in this setting.

Running with the truth preservation view of logic, a natural first suggestion is that logically valid arguments preserve truth at *all* indices in all models:

Def 2. The argument from $\varphi_1, \dots, \varphi_n$ to ψ is valid, $\{\varphi_1, \dots, \varphi_n\} \models_0 \psi$, just in case there is no model \mathcal{M} such that for some world $w \in \mathcal{W}$ and

accessibility relation $\mathcal{R} \subseteq \mathcal{W} \times \mathcal{W}$ between worlds. A recursive specification of truth at a world $w \in \mathcal{W}$ uses \mathcal{R} and \mathcal{V} to establish the complete interpretation function $\llbracket \cdot \rrbracket_{\mathcal{M}} : S_{\mathcal{L}} \times \mathcal{W} \mapsto \{T, F\}$ for this language, and $\llbracket \cdot \rrbracket_{\mathcal{M}}^w$ is obtained by holding w fixed. The semantic clauses for the informational modals, in particular, can be stated as follows:

$$\begin{aligned} \llbracket \Box \varphi \rrbracket_{\mathcal{M}}^w = T & \quad \text{iff} \quad \forall w' \in \{w' : wRw'\} (\llbracket \varphi \rrbracket_{\mathcal{M}}^{w'} = T) \\ \llbracket \Diamond \varphi \rrbracket_{\mathcal{M}}^w = T & \quad \text{iff} \quad \exists w' \in \{w' : wRw'\} (\llbracket \varphi \rrbracket_{\mathcal{M}}^{w'} = T) \end{aligned}$$

¹⁰The idea that ‘if’-clauses restrict quantificational operators goes back to Lewis [1975]. Two important asides:

(I) Yalcin [2007] requires that $i + \varphi \neq \emptyset$ but I follow Kolodny and MacFarlane [2010] and relax this restriction.

(II) If antecedent φ can include informational modal operators, then $i + \varphi$ isn’t well-defined since there needn’t be a *unique maximal φ -subset* $i' \subseteq i$ such that $\forall w \in i' (\llbracket \varphi \rrbracket_{\mathcal{M}}^{w,i'} = T)$ and $\exists w \in i'' (\llbracket \varphi \rrbracket_{\mathcal{M}}^{w,i''} = F)$ for each i'' such that $i' \subset i'' \subseteq i$. For example, when $i = \{w_1^*, w_2^*\}$, $\llbracket A \wedge \neg B \rrbracket_{\mathcal{M}}^{w_1^*, \{w_1^*\}} = T$, and $\llbracket \neg A \wedge B \rrbracket_{\mathcal{M}}^{w_2^*, \{w_2^*\}} = T$, both $\{w_1^*\}$ and $\{w_2^*\}$ are candidates for $i + \Box A \vee \Box B$.

Kolodny and MacFarlane [2010] present this semantic clause for the general case:

$$\llbracket \varphi \Rightarrow \psi \rrbracket_{\mathcal{M}}^{w,i} = T \quad \text{iff} \quad \forall i' \in i_{\varphi} (\forall w' \in i' (\llbracket \psi \rrbracket_{\mathcal{M}}^{w',i'} = T))$$

where i_{φ} is the set of maximal φ -subsets of i . But for ease of exposition, I’ll just work with the simpler semantic clause in the main body of my dissertation where I consider only indicatives with non-modal antecedents (I consider indicatives with modal antecedents in Appendix A).

information state $i \in 2^{\mathcal{W}}$, $\llbracket \varphi_1 \rrbracket_{\mathcal{M}}^{w,i} = \dots = \llbracket \varphi_n \rrbracket_{\mathcal{M}}^{w,i} = T$ and $\llbracket \psi \rrbracket_{\mathcal{M}}^{w,i} = F$.

However, Def 2 seems too demanding. This is a good argument:

(P1) Mrs. Peacock must have done it.

(C) Mrs. Peacock did it.

But $\{\Box A\} \not\models_0 A$, since when $\mathcal{V}(A, w_1^*) = T$ and $\mathcal{V}(A, w_2^*) = F$ truth is not preserved at the index $\langle w_2^*, \{w_1^*\} \rangle$. Similarly, this is good:

(P1) Mrs. Peacock did it.

(C) Mrs. Peacock might have done it.

But $\{A\} \not\models_0 \Diamond A$, since truth is not preserved at the index $\langle w_1^*, \{w_2^*\} \rangle$.

One might insist that we are simply considering too many indices. MacFarlane [2003] calls the above compositional semantics for \mathcal{L} the “semantics proper.” Importantly, the values T and F which appear in the recursive clauses are mere technical markers of help in specifying our tacit semantic competence to interpret sentences translatable into the formal language \mathcal{L} . But we can now do some “postsemantics” and define sentential *truth-at-a-context* in terms of the technical notion of *truth-at-an-index*:

Def 3. φ is true at c in \mathcal{M} if and only if $\llbracket \varphi \rrbracket_{\mathcal{M}}^{w_c, i_c} = T$

where c is the context of an actual or possible use of φ in a speech or mental act, and w_c and i_c are the world and information state supplied by c respectively.¹¹ The truth preservation conception of validity can then be reanalyzed in terms of truth at c in \mathcal{M} :

Valid_{M'} The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if there is no context c and model \mathcal{M} for the formal language \mathcal{L} such that the translations of $\varphi_1, \dots, \varphi_n$ into \mathcal{L} are all true at c in \mathcal{M} but the translation of ψ into \mathcal{L} is false at c in \mathcal{M} .

Assuming, then, that i_c explicates the knowledge of some contextually salient agent and so includes w_c given the *factivity* of knowledge, the trouble with Def 2 is that it requires truth preservation, not only at *proper* indices corresponding to contexts of use, but also at indices— $\langle w_2^*, \{w_1^*\} \rangle$, $\langle w_1^*, \{w_2^*\} \rangle$, for example—which do not correspond to an actual or possible agent’s situation and knowledge state.¹²

This suggests the following fix:

¹¹A context is typically modeled by a Lewisian *centered world* $\langle w, t, S \rangle$ consisting of a world w , time t , and agent S .

¹²cf. Kolodny and MacFarlane [2010], drawing on Kaplan [1989].

Def 4. The argument from $\varphi_1, \dots, \varphi_n$ to ψ is valid, $\{\varphi_1, \dots, \varphi_n\} \models_{Tr} \psi$ just in case there is no model \mathcal{M} such that for some information state $i \in 2^{\mathcal{W}}$ and world $w \in i$, $\llbracket \varphi_1 \rrbracket_{\mathcal{M}}^{w,i} = \dots = \llbracket \varphi_n \rrbracket_{\mathcal{M}}^{w,i} = T$ and $\llbracket \psi \rrbracket_{\mathcal{M}}^{w,i} = F$.

The relation \models_{Tr} preserves truth at all and only proper indices $\langle w, i \rangle$ where $w \in i$ in all models.¹³ So $\{\Box\varphi\} \models_{Tr} \varphi$ and $\{\varphi\} \models_{Tr} \Diamond\varphi$ for all $\varphi \in \mathcal{S}_{\mathcal{L}}$, and \models_{Tr} supports our judgments that these implications are good.

It remains to add the identity relation, quantifiers, and arguably other intensional (tense, metaphysical, etc.) operators to \mathcal{L} . But the hope, again, is that the formal relation \models_{Tr} extensionally captures the informal notion of logical validity over the fragment of our language without these expressions—that is, $\{\varphi_1, \dots, \varphi_n\} \models_{Tr} \psi$ just in case it is impossible for each of $\varphi_1, \dots, \varphi_n$ to be true at a context c and for ψ to be false at c by virtue of their logical form—here I use ‘true’ in an ordinary, pre-theoretic sense, and not in the technical sense relativized to a model \mathcal{M} . If this equivalence holds, then investigating the laws governing \models_{Tr} can teach us about the target intuitive notion and, given the purported tight link between logic and epistemology, about what we ought to believe.

¹³Since we often investigate what is so according to *non-factive* information that rules out the actual state of the world, this restriction to indices $\langle w, i \rangle$ where $w \in i$ strikes me as a bad move. However, this is not a point I want to press. I turn to other worries with \models_{Tr} in the next chapter.

Chapter 3

Against Truth Preservation

There are, however, reasons to worry about the formal relation \models_{Tr} and the truth preservation view of logic in general. Various arguments involving informational modals and the indicative conditional suggest that necessary truth preservation and good deductive argument come apart.

3.1 \models_{Tr} Versus Good Argument

First, note that \models_{Tr} also invalidates some intuitively good deductive arguments. For example, we do well to make this argument in both categorical and hypothetical deliberative contexts:

(P1) Professor Plum didn't do it.

(C) It's not the case that Professor Plum might have done it.¹

Publicly, if someone asserts or supposes that Professor Plum didn't do it, then we do well to infer on this basis that it's not the case that Professor Plum might have done it.² Privately, if you activate your belief or simply suppose in an episode of internal deliberation that Professor Plum didn't do it, then you do well to infer that it's not the case that he might have done it. But $\{\neg A\} \not\models_{Tr} \neg\Diamond A$, since when $\mathcal{V}(A, w_1^*) = T$ and $\mathcal{V}(A, w_2^*) = F$ truth is not preserved at the proper index $\langle w_2^*, \{w_1^*, w_2^*\} \rangle$ (see Fig 2).

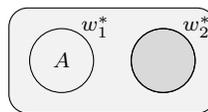


Fig 2. The index $\langle w_2^*, \{w_1^*, w_2^*\} \rangle$

¹Yalcin [2007] calls $\{\neg\varphi\} \models \neg\Diamond\varphi$ “Łukasiewicz’s Principle” since Łukasiewicz [1930] appears to endorse it.

²I use ‘infer’ in a thin sense. Inference consists of recognizing what follows; it needn’t culminate in belief.

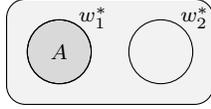


Fig 3. The index $\langle w_1^*, \{w_1^*, w_2^*\} \rangle$

Further, this argument is good:

(P1) Either Mrs. White did it or Miss Scarlett did it.

(P2) Miss Scarlett didn't do it.

(C) Mrs. White must have done it.³

But $\{A \vee B, \neg B\} \not\models_{Tr} \Box A$, since when $\llbracket A \wedge \neg B \rrbracket^{w_1^*, \{w_1^*\}} = T$ and $\mathcal{V}(A, w_2^*) = F$ truth is not preserved at the proper index $\langle w_1^*, \{w_1^*, w_2^*\} \rangle$ (see Fig 3).

Lastly, this argument is good:⁴

(P1) If a married woman committed the murder, then if Mrs. Peacock didn't do it, it was Mrs. White.

(P2) A married woman committed the murder.

(C) If Mrs. Peacock didn't do it, it was Mrs. White.

In the course of our deliberations, we do well to argue from the conditional premise P1 and its antecedent P2 to the conditional conclusion C.⁵

But $\{A \Rightarrow (\neg B \Rightarrow C), A\} \not\models_{Tr} \neg B \Rightarrow C$, since when $\llbracket A \wedge B \wedge \neg C \rrbracket^{w_1^*, \{w_1^*\}} = T$, $\llbracket A \wedge \neg B \wedge C \rrbracket^{w_2^*, \{w_2^*\}} = T$, and also $\llbracket \neg A \wedge \neg B \wedge \neg C \rrbracket^{w_3^*, \{w_3^*\}} = T$ truth is not preserved at the proper index $\langle w_1^*, \{w_1^*, w_2^*, w_3^*\} \rangle$ (see Fig 4).

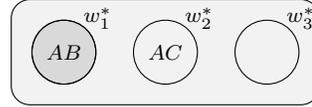


Fig 4. The index $\langle w_1^*, \{w_1^*, w_2^*, w_3^*\} \rangle$

³One might object that this argument only seems good because of the modal element in the standard characterization of logical validity—whenever the premises are true, the conclusion *must* be true—and the validity of disjunctive syllogism. But the following good argument is also invalidated by \models_{Tr} :

(P1) The murder occurred in the library and either Mrs. White did it or Miss Scarlett did it.

(P2) Miss Scarlett didn't do it.

(C) The murder occurred in the library and Mrs. White must have done it.

It is harder to hear the embedded 'must' in C as the modality of logical validity.

⁴This is a variant of McGee's [1985] famous 'counterexample' to *modus ponens* based on the 1980 U.S. Presidential election. Not everyone agrees that it is a good argument. I defend my claim in Chapter 5.

⁵This is *not* to say that in categorical deliberative contexts involving assertion and belief activation you should come to believe that if it wasn't Mrs. Peacock it was Mrs. White. Perhaps you believe that P1 holds or that P2 holds in the face of strong evidence to the contrary. Still, the *modus ponens* inference can shed important light on the normativity of your situation—for instance, that you ought, in a sense to be clarified in Chapter 6, either not to believe both that a married woman did it *and* if a married woman did it then if it wasn't Mrs. Peacock it was Mrs. White, or to believe that if it wasn't Mrs. Peacock it was Mrs. White.

The upshot: If logical validity is understood in terms of necessary truth preservation and \models_{Tr} explicates this informal target notion, then these arguments reveal that validity and good deductive argument do not line up.

3.2 Sophisticated Truth Preservation

How to respond to this mismatch? One might, of course, just dismiss my positive evaluations of the previous three arguments as misguided. However, my intuition that these are good deductive arguments is widely shared.⁶

Alternatively, one might agree that these arguments are good and broadly maintain the truth preservation view of logic, but concede that \models_{Tr} is not up to the job. In fact, the compositional semantics in Chapter 2 suggests some other consequence relations that also preserve truth at indices.

Advocates of the truth preservation view, for instance, might follow Willer [2012] and replace \models_{Tr} with this “dynamic logical consequence” relation:

Def 5. The argument from $\varphi_1, \dots, \varphi_n$ to ψ is valid, $\langle \varphi_1, \dots, \varphi_n \rangle \models_D \psi$, just in case there is no model \mathcal{M} such that for some information state $i \in 2^{\mathcal{W}}$ and world $w \in i$, $\llbracket \varphi_1 \rrbracket_{\mathcal{M}}^{w,i} = \dots = \llbracket \varphi_n \rrbracket_{\mathcal{M}}^{w,i \boxplus \dots \boxplus \varphi_{n-1}} = T$ and $\llbracket \psi \rrbracket_{\mathcal{M}}^{w,i \boxplus \dots \boxplus \varphi_{n-1} \boxplus \varphi_n} = F$.

where $i \boxplus \varphi =_{def} i \cap \{w : \llbracket \varphi \rrbracket^{w,i} = 1\}$.⁷ Informally the idea is something like this:

Valid_D The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if it is impossible, by virtue of logical form, for φ_1 to be true at context c_1 , for φ_2 to be true at context c_2 in which the salient information is the information from c_1 after being updated by φ_1 , ..., and for ψ to be false at context c_{n+1} in which the salient information is the information from c_1 after being updated sequentially by each of the premises $\varphi_1, \dots, \varphi_n$.

Logical validity, on this dynamic conception, necessarily preserves truth not at a single context c , but at a string of related contexts c_1, \dots, c_{n+1} .

⁶Again, it is not universally shared. Recall n. 4.

⁷We could define a dynamic logical consequence relation using the update $+$ from Chapter 2, but I stick to Willer’s presentation. I also use an ordered sequence of premises here because, as I’ll discuss in a moment, order matters on the dynamic conception of validity.

To its credit, \models_D validates the two good deductive arguments from Chapter 2: $\langle \Box A \rangle \models_D A$ and $\langle A \rangle \models_D \Diamond A$.⁸ Moreover, \models_D validates the three arguments that give \models_{Tr} a rough time: $\langle \neg A \rangle \models_D \neg \Diamond A$, $\langle A \vee B, \neg B \rangle \models_D \Box A$, and $\langle A \Rightarrow (\neg B \Rightarrow C), A \rangle \models_D \neg B \Rightarrow C$.⁹ Certainly, the entailment $\{A\} \models_D \Box A$ is slightly odd, but I do not think it threatens the link between \models_D and good deductive argument. I've found that many who feel the argument from 'Mrs. White did it' to 'Mrs. White must have done it' is odd also feel the argument from 'Either Mrs. White did it or Miss Scarlett did it' and 'Miss Scarlett didn't do it' to 'Mrs. White must have done it' is fine. So I suspect that the perceived oddness of the argument from A to $\Box A$ stems from its redundancy, not from its unreliability.¹⁰

However, Willer's dynamic logical consequence relation has some strange, non-standard structural properties. First, \models_D is sensitive to the order of the premises: $\langle \neg A, \Diamond A \rangle \models_D \Diamond A$ but $\langle \Diamond A, \neg A \rangle \not\models_D \Diamond A$ (hence my use of ordered pairs). Second, \models_D is not right monotonic: $\langle \Diamond A \rangle \models_D \Diamond A$ but $\langle \Diamond A, \neg A \rangle \not\models_D \Diamond A$.¹¹ It is particularly troubling that $\langle \Diamond A, \neg A \rangle \not\models_D \Diamond A$ since inferences like the following seem no more jarring than *ex falso quodlibet*:

(P1) Reverend Green might have done it.

(P2) Reverend Green didn't do it.

(C) Reverend Green might have done it.

From premises P1 and P2, it seems to me that we can infer *anything we please*—we can infer that Reverend Green might have done it, that he

⁸If $\llbracket \Box A \rrbracket_{\mathcal{M}}^{w,i} = T$ for some $w \in i$, then $\mathcal{V}(A, w) = T$, so $\llbracket A \rrbracket_{\mathcal{M}}^{w, i \boxplus \Box A} = T$. Hence $\langle \Box A \rangle \models_D A$.

If $\llbracket A \rrbracket_{\mathcal{M}}^{w,i} = T$ for some $w \in i$, then $\mathcal{V}(A, w) = T$ and $w \in i \boxplus A$, so $\llbracket \Diamond A \rrbracket_{\mathcal{M}}^{w, i \boxplus A} = T$. Hence $\langle A \rangle \models_D \Diamond A$.

⁹ $i \boxplus \neg A = \{w \in i : \mathcal{V}(w, A) = 0\}$, so $\llbracket \neg \Diamond A \rrbracket_{\mathcal{M}}^{w, i \boxplus \neg A} = T$ for each $w \in i$. Hence $\langle \neg A \rangle \models_D \neg \Diamond A$.

$(i \boxplus A \vee B) \boxplus \neg B = \{w \in i : \mathcal{V}(A, w) = 1\}$, so $\llbracket \Box A \rrbracket_{\mathcal{M}}^{w, (i \boxplus A \vee B) \boxplus \neg B} = T$ for each $w \in i$. Hence $\langle A \vee B, \neg B \rangle \models_D \Box A$.

If $\llbracket A \Rightarrow (\neg B \Rightarrow C) \rrbracket_{\mathcal{M}}^{w,i} = \llbracket A \rrbracket_{\mathcal{M}}^{w, i \boxplus A \Rightarrow (\neg B \Rightarrow C)} = T$ for some $w \in i$, then $\llbracket \neg B \Rightarrow C \rrbracket_{\mathcal{M}}^{w, i \boxplus A} = T$ and $(i \boxplus A \Rightarrow (\neg B \Rightarrow C)) \boxplus A = i \boxplus A = i + A$, so $\llbracket \neg B \Rightarrow C \rrbracket_{\mathcal{M}}^{w, (i \boxplus A \Rightarrow (\neg B \Rightarrow C)) \boxplus A} = T$. Hence $\langle A \Rightarrow (\neg B \Rightarrow C), A \rangle \models_D \neg B \Rightarrow C$.

¹⁰The consequence relation I ultimately endorse also validates the argument from A to $\Box A$. I consider a more sophisticated worry with this argument in Chapter 5.

¹¹Willer regards the non-monotonicity of \models_D as a virtue, since it allows us to “disentangle *modus ponens* from *modus tollens*” and take the former argument form to be valid and the latter to be invalid. But this is a poor motivation for \models_D since non-monotonicity is not really required to disentangle the two *modi*. More on this in n. 14 below.

must have done it, that he might not have done it, and anything else for that matter. The situation is like one where we begin reasoning from the contradictory basis that Reverend Green did it and that he didn't do it.

Admittedly, these considerations do not decisively refute the dynamic conception of logical consequence. But surely it would be preferable to retain monotonicity and the order-insensitivity of premises, if possible. Moving along, then, advocates of the truth preservation view might instead embrace this formal relation:

Def 6. The argument from $\varphi_1, \dots, \varphi_n$ to ψ is valid, $\{\varphi_1, \dots, \varphi_n\} \models_I \psi$, just in case there is no model \mathcal{M} such that for some information state $i \in 2^{\mathcal{W}}$, $\forall w \in i (\llbracket \varphi_1 \rrbracket_{\mathcal{M}}^{w,i} = \dots = \llbracket \varphi_n \rrbracket_{\mathcal{M}}^{w,i} = T)$ and $\neg \forall w \in i (\llbracket \psi \rrbracket_{\mathcal{M}}^{w,i} = T)$.

For each information state i , \models_I preserves truth across the cluster of indices $\langle w, i \rangle$ where $w \in i$: if each premise is true at each index in this cluster, then the conclusion is also true at each index in this cluster. \models_I thus motivates a kind of *global* truth preservation conception of logical validity.¹² Let C^i be the set of contexts in which the contextually salient agent's situation and knowledge state can be explicated by an index $\langle w, i \rangle$ where $w \in i$. Then one informal characterization of logical validity corresponding to \models_I is this:

Valid_{GT} The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if, for every set C^i , it is impossible, by virtue of logical form, for each of the premises $\varphi_1, \dots, \varphi_n$ to be true at every context in C^i and for the conclusion ψ to be false at some context in C^i .

Logical validity, on this global conception, necessarily preserves truth not at context c , but across the set of contexts C^i .¹³

¹²I am grateful to Arden Koehler for suggesting something like this global view.

¹³Yalcin [2007] endorses \models_I and Veltman [1996] proposes a similar consequence relation \models_3 defined over his dynamic *update semantics*. But neither Yalcin nor Veltman endorses the global truth preservation conception of logical validity that \models_I suggests. I discuss another informal analogue to \models_I (and clarify the I -subscript) in Chapter 4.

Kolodny and MacFarlane [2010] endorse both \models_{Tr} and a slight variation of \models_I . They say that the argument from $\varphi_1, \dots, \varphi_n$ to ψ is “quasi-valid” just in case there is no model \mathcal{M} such that for some information state $i \in 2^{\mathcal{W}}$ and world $w \in i$, $\llbracket \Box \varphi_1 \rrbracket_{\mathcal{M}}^{w,i} = \dots = \llbracket \Box \varphi_n \rrbracket_{\mathcal{M}}^{w,i} = T$ and $\llbracket \psi \rrbracket_{\mathcal{M}}^{w,i} = F$. Correspondingly, advocates of the truth preservation view might go for this informal conception of (quasi-)validity:

Valid_{QT} The argument from $\varphi_1, \dots, \varphi_n$ to ψ is (quasi-)valid if and only if it is impossible for each of the sentences formed by embedding each of the

Like \models_D , \models_I validates the good deductive arguments considered so far: $\{\Box A\} \models_I A$, $\{A\} \models_I \Diamond A$, $\{\neg A\} \models_I \neg \Diamond A$, $\{A \vee B, \neg B\} \models_I \Box A$, and $\{A \Rightarrow (\neg B \Rightarrow C), A\} \models_I \neg B \Rightarrow C$.¹⁴ Unlike \models_D , however, \models_I has standard structural properties; in particular, \models_I is monotonic and insensitive to the order of premises.

This is all very promising. Indeed, I'll ultimately come out in favor of the formal relation \models_I . This said, I still think that we should resist the global truth preservation view of validity that \models_I suggests. There is a deeper worry with this global view, and with the truth preservation view in general. The worry is that sentences involving informational modal operators are not, properly speaking, even true or false at a context.¹⁵

premises $\varphi_1, \dots, \varphi_n$ under an informational necessity operator to be true
and for the conclusion ψ to be false by virtue of their logical form.

But $\text{Valid}_{\Box T}$ is awkward. It is strange for the (quasi-)validity of an argument to depend on whether *another* argument necessarily preserves truth.

¹⁴If $\llbracket \Box A \rrbracket_{\mathcal{M}}^{w,i} = T$ for each $w \in i$, then clearly $\llbracket A \rrbracket_{\mathcal{M}}^{w,i} = T$ for each $w \in i$. Hence $\{\Box A\} \models_I A$.

If $\llbracket A \rrbracket_{\mathcal{M}}^{w,i} = T$ for each $w \in i$, then $\llbracket A \rrbracket_{\mathcal{M}}^{w,i} = T$ for some $w \in i$, so $\llbracket \Diamond A \rrbracket_{\mathcal{M}}^{w,i} = T$ for each $w \in i$. Hence $\{A\} \models_I \Diamond A$.

If $\llbracket \neg A \rrbracket_{\mathcal{M}}^{w,i} = T$ for each $w \in i$, then $\llbracket A \rrbracket_{\mathcal{M}}^{w,i} = F$ for each $w \in i$, so $\llbracket \neg \Diamond A \rrbracket_{\mathcal{M}}^{w,i} = T$ for each $w \in i$. Hence $\{\neg A\} \models_I \neg \Diamond A$.

If $\llbracket A \vee B \rrbracket_{\mathcal{M}}^{w,i} = \llbracket \neg B \rrbracket_{\mathcal{M}}^{w,i} = T$ for each $w \in i$, then $\llbracket A \rrbracket_{\mathcal{M}}^{w,i} = T$ for each $w \in i$, so $\llbracket \Box A \rrbracket_{\mathcal{M}}^{w,i} = T$ for each $w \in i$. Hence $\{A \vee B, \neg B\} \models_I \Box A$.

If $\llbracket A \Rightarrow (\neg B \Rightarrow C) \rrbracket_{\mathcal{M}}^{w,i} = \llbracket A \rrbracket_{\mathcal{M}}^{w,i} = T$ for each $w \in i$, then $\llbracket \neg B \Rightarrow C \rrbracket_{\mathcal{M}}^{w,i+A} = T$ for each $w \in i + A$ and $i = i + A$, so $\llbracket \neg B \Rightarrow C \rrbracket_{\mathcal{M}}^{w,i} = T$ for each $w \in i$. Hence, $\{A \Rightarrow (\neg B \Rightarrow C), A\} \models_I \neg B \Rightarrow C$.

Interestingly, \models_I validates *modus ponens* for the indicative but invalidates *modus tollens*: $\{\varphi \Rightarrow \psi, \neg \psi\} \not\models_I \neg \varphi$ for some $\varphi, \psi \in S_{\mathcal{L}}$. Consider this argument based on Lewis Carroll's [1894] barbershop paradox (cf. MacFarlane and Kolodny [2010], p. 140):

(P1) If Colonel Mustard isn't in the study, then if Professor Plum isn't in the study, Reverend Green is.

(P2) It's not the case that if Professor Plum isn't in the study, then Reverend Green is.

(C) It's not the case that Colonel Mustard isn't in the study.

$\{\neg A \Rightarrow (\neg B \Rightarrow C), \neg(\neg B \Rightarrow C)\} \not\models_I \neg \neg A$, since when $\llbracket \neg A \wedge \neg B \wedge C \rrbracket_{\mathcal{M}}^{w_1^*, \{w_1^*\}} = T$ and $\llbracket A \wedge \neg B \wedge \neg C \rrbracket_{\mathcal{M}}^{w_2^*, \{w_2^*\}} = T$, $\forall w \in \{w_1^*, w_2^*\} (\llbracket \neg A \Rightarrow (\neg B \Rightarrow C) \rrbracket_{\mathcal{M}}^{w, \{w_1^*, w_2^*\}} = T)$, $\forall w \in \{w_1^*, w_2^*\} (\llbracket \neg(\neg B \Rightarrow C) \rrbracket_{\mathcal{M}}^{w, \{w_1^*, w_2^*\}} = T)$, and $\llbracket \neg \neg A \rrbracket_{\mathcal{M}}^{w_1^*, \{w_1^*, w_2^*\}} = F$. But \models_I delivers the right verdict here since this argument is no good; we do badly to make it in categorical and hypothetical deliberative contexts. See Yalcin [2012a] for other bad *modus tollens* arguments that \models_I invalidates.

¹⁵This worry also pertains to sentences involving indicative conditionals. But I largely ignore indicatives in the remainder of this chapter.

3.3 Truth Preservation Versus Good Argument

Let me explain. Suppose that a card lies face down on a table and I tell you the following:

- (1) The card is the 4 of clubs.

Assuming that I am not just horsing around, the familiar Gricean picture is this. I believe that the card on the table is the 4 of clubs and I wish to share this information with you. I therefore assert that the card is the 4 of clubs, signaling both my belief that it is and my reflexive intention that you acquire the same belief by means of recognizing this very intention. To signal these attitudes, I have *described* our world—against the shared customs of our linguistic practice—as one in which the card on the table is the 4 of clubs. If the world is in fact this way, then I have spoken truthfully. If the card on the table is some other card in the deck, then I have spoken falsely; the world is not as I have described it.

This much is uncontroversial. At least in straightforward episodes of communication, the familiar story seems to get things right. But now suppose that I had instead uttered the following sentence in c_1^* where an informational possibility operator takes wide scope over (1):

- (2) The card might be the 4 of clubs.

Can my speech act be understood in more or less the same way as before? In particular, have I uttered a sentence whose truth at c_1^* turns on how things are in this environment? If so, what exactly have I described in using this sentence?

It is implicit in the postsemantics in Chapter 2 that informational modal talk is a kind of descriptive discourse. Recall the definition of sentential truth-at-a-context:

Def 3. φ is true at c in \mathcal{M} if and only if $\llbracket \varphi \rrbracket_{\mathcal{M}}^{w_c, i_c} = T$.

Plugging in (1) gives us:¹⁶

$4\clubsuit$ is true at c in \mathcal{M} if and only if $V(4, w_c) = V(\clubsuit, w_c) = T$.

Plugging in (2) gives us:

$\diamond 4\clubsuit$ is true at c in \mathcal{M} if and only if $\exists w \in i_c(\llbracket 4\clubsuit \rrbracket_{\mathcal{M}}^{w, i_c} = T)$.

Whereas the truth of (1) hangs on the card lying on the table, then, the truth of (2) hangs on some contextually determined information.

¹⁶Let us now assume that in addition to italicized sentence letters, \mathcal{L} includes atoms $A, 2, \dots, 10, J, Q$ and K abbreviating ‘The card is an ace’ and so forth, and atoms $\clubsuit, \diamond, \heartsuit, \spadesuit$ abbreviating ‘The card is a club’ and so forth. I omit the conjunction symbol when conjoining these new atoms—for example, $4 \wedge \clubsuit$ will be abbreviated as $4\clubsuit$.

Sentence (2) is true at context of use c_1^* just in case the salient body of information $i_{c_1^*}$ in this context rules out the possibility that the card isn't the 4 of clubs.

However, I'll now argue that this descriptivism about informational modality is suspect. I've remained silent about how context determines the information against which sentences involving informational modal operators are evaluated. Once we get down to the business of working out the contextual details, though, it is difficult, if not impossible, to specify i_c such that we can simultaneously explain informational modal talk from the first-person standpoint of deliberation, the second-person standpoint of advice, and also the third-person standpoint of appraisal.¹⁷

3.3.1 Trouble With Descriptivism

If I (agent S_1) have no idea which card is on the table in c_1^* , then I have presumably spoken felicitously in uttering (2). Why is that? Letting $\mathcal{K}_t^S(w) \in 2^{\mathcal{W}}$ designate what agent S knows at t in w , a natural speaker-centric answer is that the truth of (2) at c_1^* turns on what I know at the world $w_{c_1^*}$ and time $t_{c_1^*}$ of this context: $i_{c_1^*} = \mathcal{K}_{t_{c_1^*}}^{S_1}(w_{c_1^*})$. Since $\mathcal{K}_{t_{c_1^*}}^{S_1}(w_{c_1^*})$ leaves open the possibility that the card on the table is the 4 of clubs, $\diamond 4\clubsuit$ is true at c_1^* .¹⁸

But now suppose that you (agent S_2) know the identity of the card and so respond a moment later in c_2^* :

(3) No. The card cannot be the 4 of clubs. It's the 2 of diamonds.

On the speaker-centric, or solipsistic, contextualist line, your response is odd. After all, my assertion that the card might be the 4 of clubs is an accurate report on my knowledge state. So why do you dispute it? It is infelicitous to reject my assertion on the basis of what *you* know about the card.¹⁹

To vindicate (3), the contextualist can get fancy. She might take $i_{c_1^*}$ to model the pooled knowledge of some contextually relevant group that includes both of us, or the knowledge that such a community can acquire through certain channels of investigation (cf. Hacking [1967],

¹⁷This is well-trodden ground so I will cover it quickly. See MacFarlane [2011] for more nuanced argument.

¹⁸This speaker-centric contextualism nicely explains why Moorean sentences like 'The card might be the 4 of clubs but I know that it isn't the 4 of clubs' sound defective.

¹⁹As von Fintel and Gillies [2011] point out, the solipsistic contextualist also incorrectly predicts that this reply to (2) is fine:

(3') # OK, but I know that the card is the 2 of hearts.

Teller [1972], DeRose [1991]).²⁰ Since the truth of $\diamond 4\clubsuit$ at c_1^* then turns partially on what you know at $w_{c_1^*}$ and $t_{c_1^*}$, and your knowledge state $\mathcal{K}_{t_{c_1^*}}^{S_2}(w_{c_1^*})$ rules out the possibility that the card on the table is the 4 of clubs, you do well to disagree with me.

However, for the nonsolipsistic contextualist to also explain replies like (3) and negative evaluations of (2) from eavesdroppers and other third parties, the contextually relevant group must be massive.²¹ Indeed, it must include anyone who will ever negatively evaluate my modal claim. But then why am I willing to utter (2)? If I am ignorant about what is known, or can be known through this or that method, by this large community of assessors, then my assertion that the card might be the 4 of clubs is unwarranted.²²

The data thus push us in different directions. First-person modal discourse pushes us towards a solipsistic contextualism where i_c is the knowledge state of the speaker in c . Second-person disagreement and third-person evaluation data pushes us towards the group reading. But there seems to be no stable compromise. Make i_c subjective enough to explain the production of informational modal claims and we can no longer handle the disagreement and evaluation data. Make i_c objective enough to explain the second-person and third-person data and a gulf opens up between the circumstances under which we are prepared to make modal claims and their assertability conditions.

This is not game over for the descriptivist. von Fintel and Gillies [2011], for instance, argue that informational modals are “multiply ambiguous by design.” I actually utter (2) against a “cloud” of admissible contexts $c_{1a}^*, c_{1b}^*, c_{1c}^*, \dots$ where $i_{c_{1a}^*}, i_{c_{1b}^*}, i_{c_{1c}^*}, \dots$ explicate my knowledge at $w_{c_{1a}^*}$ and $t_{c_{1a}^*}$, your knowledge at $w_{c_{1a}^*}$ and $t_{c_{1a}^*}$, our pooled knowledge at

²⁰There are many different kinds of pooled knowledge: distributed knowledge, common knowledge, and so forth. See Fagin et al. [1995] for a good review of some of the options.

²¹As Yalcin [2011] reports, intuitions conflict in eavesdropping cases. If third parties who know that the card on the table is the 2 of diamonds are asked to evaluate (2) at c_1^* , then presumably some will say that this sentence is true, some that it is false, and others will be undecided. While I focus on the second group of negative appraisers, Yalcin argues that the contextualist has trouble explaining the conflicting intuitions themselves.

²²My assertion is unwarranted according to the popular knowledge account of assertion (Williamson [1996] and [2000]):

Knowledge Rule: You must: assert in c that φ is true only if you know in c that φ is true.

I violate this weaker rule as well:

Justified Belief Rule: You must: assert in c that φ is true only if you justifiably believe in c that φ is true.

$w_{c_{1a}^*}$ and $t_{c_{1a}^*}$, and so on. By revamping the pragmatics of assertion and denial, we can then explain informational modal talk in a contextualist manner.

Alternatively, we might abandon contextualism for a *relativism* on which sentences in $S_{\mathcal{L}}$ are true or false relative both to the context c_U in which they are uttered and to the context c_A in which they are *assessed* (cf. MacFarlane [2011]):²³

Def 7. φ is true at c_U and c_A in \mathcal{M} if and only if $\llbracket \varphi \rrbracket_{\mathcal{M}}^{w_{c_U}, i_{c_A}} = T$.

If $i_{c_1^*} = \mathcal{K}_{t_{c_1^*}}^{S_1}(w_{c_1^*})$ and $i_{c_2^*} = \mathcal{K}_{t_{c_2^*}}^{S_2}(w_{c_2^*})$, then $\diamond 4\clubsuit$ is clearly true as used at c_1^* and assessed at c_1^* but this sentence is false as used at c_1^* and assessed at c_2^* . So you are entitled to assert that the card might be the 4 of clubs and I am entitled to deny this.

However, these positions are pretty exotic. It would be nice to have a semantic theory that elucidates both the production and evaluation of informational modal claims without appealing to rampant contextual indeterminacy or assessment sensitivity—maneuvers that I, at least, find unappetizing. This brings us to expressivism.

3.3.2 Expressivism

The particular brand of expressivism that I'll discuss is developed in Yalcin [2011]. On his semantic theory, a speaker who utters a modal sentence like (2) does not describe a way the world is. This speaker does not put forward a sentence which is true or false at the context in which it is used (or true or false at each member of the set of contexts in which it is used, or true or false at the context in which it is used and the context in which it is assessed).

First question: What is it to believe in c_1^* that the card on the table might be the 4 of clubs? It is not, Yalcin argues, to have a higher-order belief about what is compatible with one's knowledge, the knowledge of some contextually relevant community of which one is a member, or some other salient body of information. Rather, letting $\mathcal{B}_t^S(w) \in 2^{\mathcal{V}}$ designate what agent S believes at t in w , and assuming that my beliefs in c_1^* are consistent, $\mathcal{B}_{t_{c_1^*}}^{S_1}(w_{c_1^*}) \neq \emptyset$, I believe that the card on the table might be the 4 of clubs just in case $\mathcal{B}_{t_{c_1^*}}^{S_1}(w_{c_1^*})$ leaves open the possibility that this is so.²⁴

²³See also Egan [2007].

²⁴Yalcin stresses that my belief state must also be sensitive to the question whether the card lying on the table is the 4 of clubs. But let me ignore this wrinkle.

It is helpful to recognize at this point that each sentence $\varphi \in S_{\mathcal{L}}$ corresponds to a possible feature $\triangleright \varphi$ of information states:

Def 8. $i \triangleright \varphi$ (read: i incorporates φ , or φ is incorporated by i) if and only if $\forall w \in i(\llbracket \varphi \rrbracket^{w,i} = T)$.²⁵

Then I believe in c_1^* that the card might be the 4 of clubs just in case $\mathcal{B}_{t_{c_1^*}}^{S_1}(w_{c_1^*}) \triangleright \diamond 4\clubsuit$.

Next question: What is it to assert in c_1^* that the card on the table might be the 4 of clubs? It is not, Yalcin argues, to describe a state of mind, a mind-independent body of information, or some other aspect of the context of use c_1^* . I do not put forward a sentence which is true or false at c_1^* . In uttering (2), I just express a certain global structural property of my state of mind—the property $\triangleright \diamond 4\clubsuit$ that my belief state shares with the belief/knowledge states of agents, or groups of agents, who believe/know that the card might be the 4 of clubs.

Situating my speech act in conversational context, my primary aim is to coordinate our states of mind on the satisfaction of $\triangleright \diamond 4\clubsuit$. Following Stalnaker [1978], Yalcin conceives of assertions as proposals to alter the *common ground* of participants in the discourses in which these acts occur. At any point in a well-run conversation, speakers presuppose the same information. They take this information for granted as the background of their conversation, or act as if they take it for granted, mutually recognizing that only certain possibilities are still on the table. The communicative impact of successful assertion is to modify this state of presupposition—as a first approximation, this can be modeled by a *context set* in $2^{\mathcal{W}}$.²⁶ If my assertion that the card on the table might be the 4 of clubs isn't rejected, then coordination is achieved: the post-assertion context set incorporates $\diamond 4\clubsuit$. If this context set explicates the informational content of our common belief, then both of our post-assertion belief states leave open the possibility that the card is the 4 of clubs.

As things would have it, though, you reject my assertion in c_2^* . Since $\mathcal{K}_{t_{c_1^*}}^{S_1}(w_{c_1^*}) \triangleright \diamond 4\clubsuit$, I am presumably entitled, in some sense, to utter (2) in c_1^* with the objective of coordinating our belief states on the openness

²⁵This correspondence is many-one—for example, $\triangleright A$ and $\triangleright \Box A$ are identical. When $i \triangleright \varphi$, Yalcin [2007] says that φ is “accepted” in i . However, I worry that ‘acceptance’ talk can suggest—at least to readers of Stalnaker [1984]—that $i \in 2^{\mathcal{W}}$ must explicate the informational content of someone’s propositional attitudes, and not information that would exist even if we did not.

²⁶See Stalnaker [2002] for a more sophisticated model of presupposition that emphasizes its public dimension. Yalcin [2012b] and [2012c] and Willer [2013] also motivate more complex conversational structure.

of $4\clubsuit$ -worlds.²⁷ However, since $\mathcal{K}_{t_{c_2}^*}^{S_2}(w_{c_2}^*) \not\models \diamond 4\clubsuit$, you ought to resist my proposal to coordinate on $\triangleright \diamond 4\clubsuit$ and, if you have a duty to inform me of the identity of the card on the table, you ought to tell me that the card is the 2 of diamonds.

Though $\mathcal{K}_{t_{c_1}^*}^{S_1}(w_{c_1}^*)$ and $\mathcal{K}_{t_{c_2}^*}^{S_2}(w_{c_2}^*)$ do not enter into the semantics or postsemantics of the sentences uttered in our discourse, what you and I know in c_2^* and c_1^* respectively still has a *pragmatic* effect, informing the assertability conditions of $\diamond 4\clubsuit$ and $\neg \diamond 4\clubsuit$. This allows Yalcin to explain the felicity of both (2) and (3) without positing complex context or assessment sensitive truth conditions. Yalcin can also easily explain negative appraisals of (2) by informed third parties who know that the card on the table is the 2 of diamonds: since the mental states of these evaluators lack $\triangleright \diamond 4\clubsuit$, they negatively assess my attempt to coordinate on this property of my mind.

Back to our main thread, though, expressivism spells trouble for the truth preservation view of logic. If sentences with informational modals are not, properly speaking, true or false, then the truth preservation view is ill-suited to explain the goodness of deductive arguments involving them. Why is the argument good from ‘The card isn’t the 4 of clubs’ to ‘It’s not the case that the card might be the 4 of clubs’? Not, according to the expressivist, because it necessarily preserves truth at each context c , across the set of contexts C^i , and so on. Since the conclusion is not truth-apt, such explanations are unsatisfactory. Indeed, if informational modal sentences are not truth-apt, then even the simple good argument from ‘The card is the 2 of diamonds’ to ‘The card might be the 2 of diamonds’ poses a problem for the truth preservation view.

Of course, much more needs to be said to motivate expressivism about informational modality over the rival contextualist and relativist positions. But the comparative ease with which expressivism explains the linguistic data counts strongly in its favor. Going forward, I assume that because of the expressivist function of informational modals, there is a disconnect between Valid_G and truth-based conceptions of logical validity like Valid_T , Valid_D , and Valid_{GT} .

3.3.3 Modality With Content²⁸

Before considering some responses to this split, I first want to defend expressivism against a pressing worry. This worry concerns the *content*

²⁷Yalcin [2011], following Gibbard [1990], calls such a speech act “rational” but “inadvisable.”

²⁸This section can be skipped without loss of continuity.

of modal sentences.

Our postsemantics in Chapter 2 consisted solely of this definition of sentential truth-at-a-context:

Def 3. φ is true at c in \mathcal{M} if and only if $\llbracket \varphi \rrbracket_{\mathcal{M}}^{w_c, i_c} = T$.

But philosophers and linguists also typically pair each sentence $\varphi \in S_{\mathcal{L}}$ in context with a set of worlds in \mathcal{W} :

Def 9. The proposition $[\varphi^c]_{\mathcal{M}}$ expressed by φ at c in \mathcal{M} is the set $\{w : \llbracket \varphi \rrbracket_{\mathcal{M}}^{w, i_c} = T\}$.

This formal object is thought to explicate the informational content of φ in c , where this content is thought to consist of the descriptivist truth condition that φ imposes on world w . Each world $w \in [\varphi^c]_{\mathcal{M}}$ designates a way the world might be that is compatible with what I describe in uttering φ in c . We can think of propositions as bearers of truth: $[\varphi^c]_{\mathcal{M}}$ is true if and only if $w_c \in [\varphi^c]_{\mathcal{M}}$.²⁹

For the expressivist, however, this kind of postsemantics founders. The recursive specification of truth-at-a-point in Chapter 2 determines the extension $\llbracket \varphi \rrbracket^{w, i}$ of each $\varphi \in S_{\mathcal{L}}$ but the information state parameter i in the index $\langle w, i \rangle$ is what Yalcin calls a “nonfactualist parameter” that isn’t initialized by a context of use c —there is no such thing as the information state i_c of this context. Thus, the truth of sentence φ at c in \mathcal{M} and the proposition $[\varphi^c]_{\mathcal{M}}$ expressed by φ at c in \mathcal{M} aren’t well-defined.

If φ is a non-modal sentence, of course, then the information state parameter is inert: $\llbracket \varphi \rrbracket^{w, i}$ does not depend on the value of i . So the expressivist can redefine truth-at-a-context and the proposition $[\varphi]_{\mathcal{M}}$ as follows:

Def 10. For non-modal sentence φ , φ is true at c in \mathcal{M} if and only if $\forall i \in 2^{\mathcal{W}} (\llbracket \varphi \rrbracket_{\mathcal{M}}^{w_c, i} = T)$.³⁰

Def 11. For non-modal sentence φ , the proposition $[\varphi]_{\mathcal{M}}$ expressed by φ at c in \mathcal{M} is the set $\{w : \forall i \in 2^{\mathcal{W}} (\llbracket \varphi \rrbracket_{\mathcal{M}}^{w, i} = T)\}$.³¹

For modal sentences, these notions are left undefined.

²⁹Philosophers and linguists commonly regard propositions as the compositional semantic values of declarative sentences—that is, they take propositions to be part of the *semantics proper*. However, Lewis [1980] stresses that semantic value and content needn’t coincide; it is enough that we can systematically recover a sentence’s informational content in context from its semantic value together with a specification of this context.

³⁰Existential quantification over $2^{\mathcal{W}}$ would also do.

³¹Since I’m ignoring the role of context in fixing the denotation of ‘the card,’ $[\varphi]_{\mathcal{M}}$ does not depend on context and I omit the c -superscript.

But now the worry. If the informational content of φ is modeled by $[\varphi]_{\mathcal{M}}$, then the revised postsemantics suggests that informational modal sentences do not express informational content. However, \Box -claims seem contentful since agents who make them can rule out possible states of the world. If you successfully assert that the card is the queen of hearts, then the context set of our conversation comes to exclude $\neg Q\heartsuit$ -worlds. Likewise if you successfully assert that the card on the table *must* be the queen of hearts. Indeed, since an agent who utters $Q\heartsuit$ and $\Box Q\heartsuit$ eliminates the same possibilities, these two sentences presumably express the *same* informational content.³²

It seems counterintuitive to deny that some informational modal claims are contentful. But in fact the expressivist needn't deny this. The key move is simply to reject the equation of informational content and descriptivist truth conditions.

Here is Stalnaker [1981] on propositions:

Propositions, according to the account I will sketch, are functions from possible worlds into truth-values. Equivalently, but more informally, they are ways of dividing a space of possibilities—ways of picking out some subset from a set of alternative ways that things might be. The intuitive idea behind this conception of proposition is an old one: it is the idea that what is essential to propositional or informational content is that certain possibilities be excluded. To say or believe that some of the ways the world might have been are not ways that it is. The *content* of what one says or believes should be understood in terms of the possibilities that are excluded. (p. 134)

³²In saying this, I am in good company. Frege suggests in the *Begriffsschrift* [1879] that your modal *apodictic* judgment differs from your non-modal *assertoric* judgment only in what you indicate about your evidential basis for making it. In asserting that the card is the queen of hearts, you signal your knowledge of general laws from which it follows that the card is the queen of hearts. In asserting that the card must be the queen of hearts, you do no such thing:

What distinguishes the apodeictic from the assertoric judgment is that it indicates the existence of general judgments from which the proposition may be inferred—an indication that is absent in the assertoric judgment. If I term a proposition 'necessary,' then I am giving a hint as to my grounds for judgment. *But this does not affect the conceptual content of the judgment; and therefore the apodeictic form of a judgment has not for our purposes any significance.* (p. 4)

Your judgments are otherwise identical. In both cases, you describe a particular way the world is—namely, one in which the card on the table is the queen of hearts. You put forward a sentence whose truth turns on which card is lying on the table.

Despite denying that informational modal sentences serve to represent reality, the expressivist can still, I submit, hold onto the core ‘ruling out’ conception of content. He can still adopt the intuitive Stalnakerian idea that a sentence has content by virtue of it excluding possibilities. Non-modal sentences carve up the space of alternative possibilities into those that conform and those that fail to conform to what is described. Informational modal sentences can also carve up this space, albeit in a less direct way—for instance, $\Box Q\heartsuit$ lacks descriptivist truth conditions but nevertheless rules out $\neg Q\heartsuit$ -worlds since these are not included in any information state that instantiates $\triangleright \Box Q\heartsuit$. Hence these modal sentences have content.

Formally, the expressivist can define the *w-eliminative content* of φ as follows:³³

Def 12. The *w-eliminative content* $[\varphi]_{\mathcal{M}}^w$ carried by φ in \mathcal{M} is the set $\bigcup\{i \in 2^{\mathcal{W}} : i \triangleright \varphi\}$.

$w \in [\varphi]_{\mathcal{M}}^w$ if and only if w is a member of some information state i that incorporates φ : $[\varphi]_{\mathcal{M}}^w = \{w : \exists i \in 2^{\mathcal{W}}(w \in i \wedge i \triangleright \varphi)\}$. So each non-member $w \notin [\varphi]_{\mathcal{M}}^w$ that is a live option in a conversation is ruled out if one of the participants utters φ and the context set comes to incorporate this sentence.

The *w-eliminative content* of a non-modal sentence φ is the proposition it expresses: $[\varphi]_{\mathcal{M}}^w = \bigcup\{i : i \triangleright \varphi\} = \{w : \forall i([\varphi]_{\mathcal{M}}^{w,i} = T)\} = [\varphi]_{\mathcal{M}}$.³⁴ For instance, $[Q\heartsuit]_{\mathcal{M}}^w = \{w : \mathcal{V}(Q, w) = \mathcal{V}(\heartsuit, w) = T\}$. However, the definition of *w-eliminative content* also applies to informational modal sentences and delivers pleasing results over the modal fragment of $S_{\mathcal{L}}$. For instance, $[\Box Q\heartsuit]_{\mathcal{M}}^w = \bigcup\{i : i \triangleright \Box Q\heartsuit\} = [Q\heartsuit]_{\mathcal{M}}^w$. Note that $Q\heartsuit$ and $\Box Q\heartsuit$ have the same *w-eliminative content* but divide up \mathcal{W} in different ways: $Q\heartsuit$ directly imposes a descriptivist constraint on worlds, whereas $\Box Q\heartsuit$ indirectly imposes a constraint on worlds via a global constraint on information states.³⁵

Interestingly, $[\varphi]_{\mathcal{M}}^w \neq \mathcal{W} \setminus [\neg\varphi]_{\mathcal{M}}^w$ for some φ and \mathcal{M} . If \mathcal{W} includes both a $Q\heartsuit$ -world and a $\neg Q\heartsuit$ -world, $\mathcal{W} \setminus [\neg\Diamond Q\heartsuit]_{\mathcal{M}}^w = [Q\heartsuit]_{\mathcal{M}}^w \neq \mathcal{W}$ but $[\Diamond Q\heartsuit]_{\mathcal{M}}^w = \mathcal{W}$. So $\neg\Diamond Q\heartsuit$ has *w-eliminative content* whereas $\Diamond Q\heartsuit$ does

³³I do not use the term ‘proposition’ because of its descriptivist connotations. I do not use the term ‘informational content’ because the *w-eliminative content* of φ in \mathcal{M} models only one dimension of informational content. More on this in n. 37.

³⁴For non-modal φ , if $w \in \bigcup\{i : i \triangleright \varphi\}$, then $[\varphi]_{\mathcal{M}}^{w,i} = T$ for some i where $w \in i$, so $\forall i \in 2^{\mathcal{W}}([\varphi]_{\mathcal{M}}^{w,i} = T)$ since $[\varphi]_{\mathcal{M}}^{w,i}$ does not depend on the value of i . Conversely, if $\forall i \in 2^{\mathcal{W}}([\varphi]_{\mathcal{M}}^{w,i} = T)$, then $[\varphi]_{\mathcal{M}}^{w,\{w\}} = T$, so $w \in \bigcup\{i : i \triangleright \varphi\}$.

³⁵Another pleasing result: $[A \Rightarrow \spadesuit]_{\mathcal{M}}^w = \bigcup\{i : i \triangleright A \Rightarrow \spadesuit\} = [A \supset \spadesuit]_{\mathcal{M}}^w$. In general, indicative and material conditionals with the same atomic antecedent and consequent have the same *w-eliminative content*.

not. This honors the nondescriptivist idea that to say that something might be the case is not to rule out any possible states of affairs. It is rather to project one’s ignorance about how things are.³⁶

Admittedly, $[\varphi]_{\mathcal{M}}^w$ does not capture the full communicative import of assertion. As stressed, sentences like $Q\heartsuit$ and $\Box Q\heartsuit$ with different syntactic forms can rule out the same worlds in \mathcal{W} in very different ways. Sentences like $\Diamond Q\heartsuit$ and $Q\heartsuit \vee \neg Q\heartsuit$ with the same w -eliminative content can also call for very different coordination of our mental states.³⁷ Still, I hope to have shown how to be an expressivist about informational modality—and so how elegantly to handle the linguistic data in §3.3.1—without implausibly committing oneself to the idea that informational modal sentences with the potential to eliminate options lack content.

³⁶Frege suggests something like this treatment of \Diamond -claims in the *Begriffsschrift*:

If a proposition is presented as possible, then either the speaker is refraining from judgment, and indicating at the same time that he is not acquainted with any laws from which the negation of the proposition would follow; or else he is saying that the negation of the proposition is in general false. In the latter case, we have what is usually termed a *particular affirmative judgment*. ‘It is possible that the Earth will one day collide with another celestial body’ is an example of the first case; ‘a chill may result in death,’ of the second case. (p. 5)

If *judging* is uttering a sentence with informational content, then Frege suggests here that informational possibility claims like ‘It is possible that the Earth will one day collide with another celestial body’ do not carry informational content. To make such a claim is not to distinguish among the possible worlds that are live options in the context of use, but is only to signal a lack of knowledge of laws from which the negation of the prejacent follows. (The ‘may’ in ‘a chill may result in death’ is a *circumstantial* modal, not an epistemic modal, so Frege’s second case is irrelevant to our discussion.)

³⁷ $[\Diamond Q\heartsuit]_{\mathcal{M}}^w = [Q\heartsuit \vee \neg Q\heartsuit]_{\mathcal{M}}^w = \mathcal{W}$. To distinguish the communicative import of these sentences, we might model the context set of a conversation by a *set* of information states $I \subseteq 2^{\mathcal{W}}$ (cf. Willer [2013]). We can then define the *i-eliminative content* of φ in \mathcal{M} as follows:

$$[\varphi]_{\mathcal{M}}^i = \{i \in 2^{\mathcal{W}} : i \triangleright \varphi\}$$

$$[Q\heartsuit \vee \neg Q\heartsuit]_{\mathcal{M}}^i = 2^{\mathcal{W}} \text{ but } [\Diamond Q\heartsuit]_{\mathcal{M}}^i \neq 2^{\mathcal{W}} \text{ so long as there is some } i \in 2^{\mathcal{W}} \text{ such that } i \not\triangleright \Diamond Q\heartsuit.$$

Chapter 4

The Informational View of Logic

Expressivism spells trouble for the truth preservation view of logic. If sentences involving informational modal operators are not truth-apt, then there is a tension between these two standard characterizations of logical validity:

Valid_T The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if it is impossible for each of the premises $\varphi_1, \dots, \varphi_n$ to be true and for the conclusion ψ to be false by virtue of their logical form.

Valid_G The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if it is a *good* deductive argument that we do well to make in both categorical and hypothetical deliberative contexts by virtue of its logical form.

At this point, one might accept the separation of truth preservation and good deductive argument and come out on one or the other side of the fence.

4.1 Responses

This response might already have come to mind:

“Why insist that all good deductive arguments are logically valid? Robert Stalnaker [1975] famously distinguishes between the semantic concept of *entailment* and the pragmatic concept of *reasonable inference*. The *or-to-if* inference from ‘Either the butler did it or the gardener did it’ to ‘If the butler didn’t do it, then the gardener did,’ says Stalnaker, is logically invalid since the former sentence does not even semantically entail the latter, but it is still a reasonable inference. Similarly, one might say that

the inference from ‘The butler didn’t do it’ to ‘It’s not the case that the butler might have done it’ is logically invalid but still a good inference.”¹

Of course, we might just stipulate ‘logically valid’ is to mean something like ‘necessarily truth preserving by virtue of logical form’ and accept that many good deductive arguments fall outside its extension. However, this move is unattractive. Though the idea that logically valid arguments necessarily preserve truth by virtue of logical form is well entrenched in the philosophical tradition, so too is the idea that good deductive arguments are logically valid. We maintain one of these ideas only by seriously undermining another.

This response might have also come to mind:

“Why not surrender the idea that logically valid arguments preserve truth? Perhaps we can understand logic in some other way such that validity and good deductive argument coincide.”

Hartry Field [2006], [2008], [2009a], [2009b], [*ms.*], for one, has argued for this approach in recent work.² He [2009a] boldly suggests that we

¹While I bring up Stalnaker’s distinction between entailment and reasonable inference to stimulate this response, I should emphasize that, as I understand these concepts, good deductive inference and reasonable inference do not extensionally coincide. Inspired by Stalnaker’s definition, let us call an inference from $\varphi_1, \dots, \varphi_n$ to ψ a *reasonable inference* iff there is no model \mathcal{M} such that for some $i \in 2^{\mathcal{W}}$, one can appropriately assert φ_1 against background information i , ..., and also appropriately assert φ_n against $i + \varphi_1 + \dots + \varphi_{n-1}$, but $i + \varphi_1 + \dots + \varphi_n \not\vdash \psi$. All the good inferences from Chapter 3 are reasonable given the semantics under consideration. However, while the *or-to-if* inference from $A \vee \diamond A$ to $\neg A \Rightarrow \diamond A$ is reasonable given the Gricean assumption that one can appropriately assert a disjunction $\varphi \vee \psi$ against background information i only if $\exists w \in i([\varphi \wedge \neg \psi]^{w,i} = T)$ and $\exists w \in i([\neg \varphi \wedge \psi]^{w,i} = T)$ (since it’s *never* appropriate to assert $A \vee \diamond A$), this inference is not good; one does poorly to make it in both categorical and hypothetical contexts.

²Field thinks that we should abandon the truth preservation view not for the reasons discussed in Chapter 3, but because of various inferences involving a general untyped truth predicate $Tr(x)$. If we look at our best formal truth theories developed since the 1970s to handle the Liar paradox and related semantic paradoxes, Field argues, these theories formulated in a language with $Tr(x)$ include axioms they do not regard as true, or rules of inference they do not regard as unrestrictedly truth preserving. Worse, adding to truth theory T either the sentence saying that all of T ’s axioms are true or the sentence saying that all of T ’s rules of inference preserve truth results in inconsistency. But all of us accept, or should accept, one or the other of these theories, taking its axioms/rules to govern our inferential practices. So if we want to hold onto the idea that logic—where ‘logic’ is broad enough to include the logic of $Tr(x)$ —lines up with good deductive inference such that these axioms/rules are logical truths/logically valid, then we are not in a position to consistently accept that logic is about unrestricted truth preservation.

In this essay, however, I am less concerned with Field’s truth-theoretic argument than with his normative characterization of logic (I’ll say a bit more about Field’s

should regard the normative dimension of logic as fundamental:

[I]f logic is not the science of what [forms of inference] necessarily preserve truth, it is hard to see what the subject of logic could possibly be, if it isn't somehow connected to norms of thought.
(p. 263)

This is not to say that logical validity should be *defined* in terms of its normativity for thought. Field finds this tack repugnant, and argues that it is best not to define logical validity at all but to treat it as a primitive notion and illuminate its conceptual role. Field [*ms.*] suggests the following role for validity:

To regard an inference or argument as valid is (in large part anyway) to accept a constraint on belief: one that prohibits fully believing its premises without fully believing its conclusion. (The prohibition should be 'due to logical form': for any other argument of that form, the constraint should also prohibit fully believing the premises without fully believing the conclusion.)
(p. 11)³

Moreover, there is nothing more to understanding logical validity than understanding what it is to regard an argument as valid. If you and I disagree over whether *modus ponens* for the indicative is valid, then one of us accepts a constraint on belief that the other rejects—that one ought either not to fully believe in φ , not to fully believe in $\varphi \Rightarrow \psi$, or to fully believe in ψ . But there needn't be any ultimate metaphysical basis for accepting one view over the other. In the course of our dispute, of course, you or I might appeal to certain objective facts to back up a position. You might argue for the belief constraint, say, by explaining that if φ and $\varphi \Rightarrow \psi$ are both true, then ψ is true as well. But the discussion in Chapter 3 shows that this kind of justification is not always available.

argument in Chapter 7). As discussed, there are reasons aside from the semantic paradoxes to think that necessary truth preservation and good deductive argument come apart.

³This proposal is refined in Field [2009a], §I, and [*ms.*], §2. He ultimately plumps for this constraint on conditional degrees of belief:

If $A_1, \dots, A_n \Rightarrow B_1, \dots, B_m$, then $\sum_{i \leq n} Dis(A_i|C/D) + \sum_{j \leq m} Cr(B_j|C/D) \geq 1$

for any C and D , where $A_1, \dots, A_n \Rightarrow B_1, \dots, B_m$ is a sequent, $Dis(A_i|C/D)$ is one's degree of disbelief in A_i conditional on full acceptance of C and full rejection of D , and $Cr(B_j|C/D)$ is one's credence in B_j conditional on full acceptance of C and full rejection of D . But for present purposes, this level of detail is unimportant; I'll just focus on the cruder proposal stated in the main text.

When it is not, our differing judgments about validity will simply reflect the different norms that we take to govern our epistemic practices.⁴

However, this Fieldian position also involves radically abandoning orthodoxy. The arguments in Chapter 3 invite us to revisit what we talk about when we talk about logic. But, once we do so, it's not obvious why we should conclude as Field [2009a] does that “logic is essentially normative” (p. 268). Not only must we give up the standard truth preservation view of logic, we must give up the more basic idea that logic is a descriptive science. This seems drastic. Is there a better option?

4.2 A Better Option

There is an alternative conception of logic that Field overlooks. Like the truth preservation view, this neglected alternative takes logic to be a descriptive science—indeed, it can help us understand the widespread attraction of truth preservation. But on this alternative, validity also coincides with good deductive argument.

Recall this formal consequence relation from §3.2:

Def 6. The argument from $\varphi_1, \dots, \varphi_n$ to ψ is valid, $\{\varphi_1, \dots, \varphi_n\} \models_I \psi$, just in case there is no model \mathcal{M} such that for some information state $i \in 2^{\mathcal{W}}$, $\forall w \in i (\llbracket \varphi_1 \rrbracket_{\mathcal{M}}^{w,i} = \dots = \llbracket \varphi_n \rrbracket_{\mathcal{M}}^{w,i} = T)$ and $\neg \forall w \in i (\llbracket \psi \rrbracket_{\mathcal{M}}^{w,i} = T)$.

\models_I delivers the right verdicts for arguments involving informational modals and the indicative. However, \models_I suggests an informal global truth preservation conception of validity that comes under fire from the expressivist about informational modality.

Let me now point out, though, that the relation \models_I has another informal analogue.

Def 6 can be restated using the notion of *incorporation* from §3.3:

$\{\varphi_1, \dots, \varphi_n\} \models_I \psi$ just in case there is no model \mathcal{M} such that for some $i \in 2^{\mathcal{W}}$, $i \triangleright \varphi_1, \dots, i \triangleright \varphi_n$, and $i \not\triangleright \psi$.

So facts about \models_I —what I'll now, following Yalcin [2007], call ‘informational consequence’—correspond to facts about information states. We can think of informational consequence as preserving not truth at indices, but incorporation in all information states: \models_I holds between a set of premises $\{\varphi_1, \dots, \varphi_n\}$ and a conclusion sentence ψ just in case any information state with the structural features $\triangleright \varphi_1, \dots, \triangleright \varphi_n$ determined by each of the premises (consisting entirely of A -worlds, containing at

⁴Field [*ms.*], §3, motivates this ‘projectivism’ about logical validity by analogy to the concept of chance.

least one B -world, and so on) has the structural feature $\triangleright \psi$ determined by the conclusion.

Of course, as we should distinguish the technical notions of truth-at-an-index and truth-at-a-context (applicable to non-modal sentences) from the ordinary pre-theoretic notion of truth, we should distinguish the technical notion of incorporation relating a sentence $\varphi \in S_{\mathcal{L}}$ and a mathematical object $i \in 2^{\mathcal{W}}$ from the ordinary pre-theoretic sense in which a body of information incorporates, say, that Professor Plum might have done it. That is, we should distinguish the theoretical fact that the sentence ‘Professor Plum might have done it’ is incorporated by information state i from the pre-theoretic fact that a particular body of information is information according to which Professor Plum might have done it. So \models_I suggests this informal characterization of logical validity:

Valid_I The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if any body of information with the structural features corresponding to each of the premises $\varphi_1, \dots, \varphi_n$ also has the structural feature corresponding to the conclusion ψ by virtue of the logical form of these sentences.⁵

Why is the argument logically valid from ‘Professor Plum didn’t do it’ to ‘It’s not the case that Professor Plum might have done it’? \models_I suggests: the argument is valid because information has a particular kind of *structure*. Information incorporating that Professor Plum didn’t do it rules out the possibility that he did it, and is therefore also information incorporating that it’s not the case that Professor Plum might have done it. Crucially, the suggestion is *not* that this argument is valid because it necessarily preserves truth (at a context or set of contexts). It’s not: ‘Professor Plum didn’t do it’ logically implies ‘It’s not the case that Professor Plum might have done it’ because it is impossible for the former sentence to be true and for the latter to be false by virtue of their logical form. It’s instead: ‘Professor Plum didn’t do it’ logically implies ‘It’s not the case that Professor Plum might have done it’ because any body of information (the content of an eyewitness’s utterances, the evening news, and so forth) according to which Professor Plum didn’t do it is therefore, by virtue of logical form, also information according to which it’s not the case that he might have done it.⁶

⁵Again, I’m using ‘information’ here in the rough, intuitive sense of that which eliminates certain possibilities while leaving others open, and not in the technical sense of a set of possible worlds $i \in 2^{\mathcal{W}}$.

⁶This suggestion also applies to arguments validated by \models_{Tr} . ‘Mrs. Peacock did

Note that we have, in both §3.2 and this section, been turning the study of logic on its head. The core concern of mathematical logic is commonly taken to be the formulation of a formal concept of logical validity that extensionally coincides with the informal concept cashed out in terms of necessary truth preservation. But formal consequence relations have spurred us to reconsider the informal concept of logical validity itself. Underlying the distinction between \models_{Tr} and \models_I is, I'm suggesting, a deeper distinction between two different ways of thinking about the target informal notion of logical validity: there is the standard truth preservation view of logic, and also what I'll call the 'informational view' on which logic is fundamentally concerned with the structure of information.⁷

Fortunately, the informational view allows us to make sense of most of the things that have been said about logic over the years. We can still regard logic as a descriptive science. We can even make sense of the considerable popularity of the truth preservation view itself. Restricted to sentences in $S_{\mathcal{L}}$ without informational modals and the indicative, $\{\varphi_1, \dots, \varphi_n\} \models_{Tr} \psi$ if and only if $\{\varphi_1, \dots, \varphi_n\} \models_I \psi$.⁸ So it is little wonder on the informational view that truth preservation has been so influential.

Further, we can maintain the idea that logical validity and good deductive argument coincide. A deductive argument counts as valid on the informational view if and only if we do well to make it in both categorical and hypothetical deliberative contexts by virtue of logical form—or so I'll argue. Let me now respond to some threats to this equivalence.

it' logically implies 'Mrs. Peacock might have done it' because information according to which Mrs. Peacock did it is therefore also information according to which she might have done it.

⁷You might think of the informal notion of incorporation as the core concern of logic understood informationally: logically valid arguments preserve incorporation in bodies of information, logical truths are incorporated in all bodies of information, logically consistent sentences are jointly incorporated in some body of information, and so forth.

⁸The left-to-right direction clearly holds for the full language. The right-to-left direction holds for sentences without informational constants since $\llbracket \varphi \rrbracket^{w,i} = T$ iff $\{w\} \triangleright \varphi$ for such sentences.

Chapter 5

Modus Ponens Defended

The claim that Valid_G and Valid_I coincide is controversial. Is it really the case that we do well to employ *modus ponens* for the indicative conditional and other argument forms that are validated by \models_I but not by \models_{Tr} in both categorical and hypothetical contexts by virtue of logical form?¹ Kolodny and MacFarlane [2010] and Willer [2010], for instance, argue that *modus ponens* for the indicative is unreliable in some deliberative contexts. As an illustrative example, I now want to defend *modus ponens* against these attacks.²

5.1 Kolodny and MacFarlane

Kolodny and MacFarlane [2010] are, in a sense, pluralists about logic. As mentioned earlier in Chapter 3, n. 13, they endorse both \models_{Tr} and a variation of \models_I . They claim that while argument forms like *modus ponens* that are validated by the latter but not by the former relation are good to make in categorical but not hypothetical contexts, \models_{Tr} tracks good argument in both categorical *and* hypothetical contexts. However, if we have a reason to care about the relation \models_{Tr} , I submit that this is not it. Again, I'll argue that informational consequence, not \models_{Tr} , lines up nicely with good deductive argument across all deliberative contexts.

5.1.1 *Reductio*

Suppose you have been in your office for hours with the blinds down and have not heard the weather forecast. Given your evidence, you reflect

¹It is easy to verify that *modus ponens* for the indicative is logically valid on the informational view: if $i \triangleright \varphi \Rightarrow \psi$ and $i \triangleright \varphi$, then $i + \varphi \triangleright \psi$ and $i = i + \varphi$, so $i \triangleright \psi$.

²Much of what I say in this chapter generalizes to other potentially problematic inferences for the informational view of logic. For instance, I discuss the inference from φ to $\Box\varphi$ in n. 9 below.

that the streets might not be wet, but that if it's raining the streets must be wet. You then enter a hypothetical deliberative context by supposing that it's raining. Applying *modus ponens* inside this context, you infer that the streets must therefore be wet. Recognizing that this conflicts with the first premise, you then conclude by *reductio ad absurdum* that it's not raining. Something has gone horribly wrong.

| | | |
|---|--|-------------|
| 1 | The streets might not be wet | Premise |
| 2 | If it's raining, the streets must be wet | Premise |
| 3 | It's raining | Supposition |
| 4 | The streets must be wet | From 2,3 |
| 5 | It's not the case that the streets must be wet | From 1 |
| 6 | \perp | From 4,5 |
| 7 | It's not raining | From 3,4-6 |

Kolodny and MacFarlane pinpoint the use of *modus ponens* at step 4 as the source of the trouble. If they are right, then this argument form tells against the informational view of logic. It seems that the informational view overgenerates by counting bad deductive inferences as logically valid.

However, is *modus ponens* really unreliable inside the hypothetical context? It seems to me that Kolodny and MacFarlane misdiagnose the problem, which arises only at step 5 or step 7, depending on how your supposition works at step 3. Consider the *informational background* of your deliberation. Sitting in the office, your information leaves open the possibility that the streets aren't wet, but rules out that it's raining and the streets aren't wet. If your supposition consists of tentatively *adding* the information that it's raining to your basic information, then your nondegenerate information in the induced hypothetical context rules out the possibility that the streets aren't wet, so you shouldn't reflect at step 5 that it's not the case that the streets must be wet.³

³The root of Kolodny and MacFarlane's misdiagnosis, on this proposal, is their assimilation of good deductive argumentation to argumentation according to a natural deduction proof system that allows unrestricted use of premises within a subproof. Though he does not discuss this particular example, Willer [2012] would also pin the blame on step 5. He argues that we must restrict what can be used inside hypothetical contexts when working with informational modals and the indicative because logical consequence is non-monotonic (recall his dynamic relation \models_D from §3.2). Note that $\langle \diamond \neg W, R \Rightarrow \Box W \rangle \models_D \neg \Box W$, but $\langle \diamond \neg W, R \Rightarrow \Box W, R \rangle \not\models_D \neg \Box W$. Thus, while you can infer in the main categorical context of your deliberation that it's not the

On the other hand, your supposition might trigger a hypothetical context in which your salient body of information rules out that it's not raining but also still has the structural features corresponding to both premises.⁴ In this case, you do well to recognize the contradiction. But while your information in the hypothetical context is degenerate—it can be explicated by the empty set \emptyset —you *cannot* conclude from this that it's not raining. It follows only that your actual information doesn't rule out that it's not raining. That is, you can conclude that it *might* not be raining. Going forward, I assume that your supposition works in this second way, so the problem with the *reductio* is just that your conclusion is overly strong.

In a close variation of this example that also involves the use of *modus ponens* in a hypothetical context, you *can* conclude that it's not raining. Suppose instead that the blinds in your office are raised high enough to see the streets but not the sky. Given your evidence, you reflect that the streets aren't wet, but that if it's raining the streets must be wet. In fact, your evidence rules out that it's raining, but you're slow to realize this so you suppose that it's raining. Applying *modus ponens* in this hypothetical context, you infer that the streets must be wet, and therefore that they're wet. Recognizing that this conflicts with the first premise, you conclude by *reductio* that it's not raining.

case that the streets must be wet, you cannot infer this in the hypothetical context triggered by the supposition that it's raining.

However, unlike Willer, I'm suggesting here that a subproof restriction is violated in the *reductio* proof not because of the non-monotonicity of logical consequence, but because of how your supposition that it's raining affects the informational background of your deliberation.

⁴In Appendix A, I distinguish this *lossless* kind of supposition from the earlier *lossy* kind. Lossy supposition triggers hypothetical contexts in which the premises of an argument can fail to hold. Lossless supposition always triggers hypothetical contexts in which one's information incorporates everything that was incorporated before.

| | | |
|---|--|-------------|
| 1 | The streets aren't wet | Premise |
| 2 | If it's raining, the streets must be wet | Premise |
| 3 | It's raining | Supposition |
| 4 | The streets must be wet | From 2,3 |
| 5 | The streets are wet | From 4 |
| 6 | The streets aren't wet | From 1 |
| 7 | \perp | From 5,6 |
| 8 | It's not raining | From 3,4-7 |

The only significant difference between this episode of reasoning and the previous one is that the first premise is now non-modal.⁵ But your argumentation, though protracted, is now impeccable. This strongly suggests that the problem with the previous example was not the use of *modus ponens*. To be fair, Kolodny and MacFarlane claim only that *modus ponens* will sometimes lead you astray in hypothetical contexts, not that it always will. However, this does not satisfactorily explain why you can argue to the stronger conclusion that it's not raining in this second example, but not in the first example.

Here is an explanation that leaves *modus ponens* untouched. If you were to gain new evidence in this second example that ruled out some possibilities left open by your original evidence, then you could still reflect that the streets aren't wet, but that if it's raining the streets must be wet. Unlike the modal premise in the first example, both premises in this second example continue to hold when the informational background of your deliberation contracts—and in particular, when it contracts to any single possibility where it's raining. The conflict between these two premises and your supposition that it's raining thus reveals not just that your original information doesn't rule out that it's not raining, but also that this information rules out the possibility that it's raining. On the basis of this conflict, you can felicitously conclude that it's not raining.⁶

⁵The second example also involves the inference from 'The streets must be wet' to 'The streets are wet' but this inference is uncontroversially valid.

⁶Some formalism can help sharpen this point. Let i^* explicate your information. It rules out the possibility that the streets are wet, and rules out the possibility that it's raining and the streets aren't wet, so $i^* \triangleright \neg W$ and $i^* \triangleright R \Rightarrow \Box W$. Consider some arbitrary $w^* \in i^*$. Since $\{w^*\} \triangleright \neg W$ and $\{w^*\} \triangleright R \Rightarrow \Box W$, $\mathcal{V}(R, w^*) = T$ only if $\{w^*\} \triangleright \perp$. Hence $\mathcal{V}(R, w^*) = F$, and since w^* was arbitrary, $i^* \triangleright \neg R$.

The real lesson from the above examples is not that *modus ponens* for the indicative is unreliable in hypothetical contexts, but instead that we must exercise caution when using *reductio ad absurdum* in languages with informational modals and the indicative. These languages include *non-persistent* sentences (cf. Veltman [1996]).

Def 13. Sentence $\varphi \in S_{\mathcal{L}}$ is persistent if and only if there is no model \mathcal{M} such that for some $i, i' \in 2^{\mathcal{W}}$ where $i' \subset i$, $i \triangleright \varphi$ and $i' \not\triangleright \varphi$.

$\neg W$, $R \Rightarrow \Box W$, and R are all persistent, but $\Diamond \neg W$ is non-persistent since when $\mathcal{V}(W, w_1^*) = T$ and $\mathcal{V}(W, w_2^*) = F$, $\{w_1^*, w_2^*\} \triangleright \Diamond \neg W$ and $\{w_1^*\} \not\triangleright \Diamond \neg W$. We should distinguish between these different forms of indirect proof:⁷

1-*reductio*: If \perp follows from A and the premise set Γ , then $\Diamond \neg A$ follows from Γ .

2-*reductio*: If \perp follows from A and only *persistent* members of Γ , then $\neg A$ follows from Γ .

2-reductio is used in the second example. But the *reductio* in the first example fits neither of these good forms.⁸ (See Appendix B for a proof system, **Info**, that is sensitive to the distinction between *1-reductio* and *2-reductio*.)

5.1.2 Constructive Dilemma

Now, one might also worry that *modus ponens* for the indicative can lead a reasoner astray in a different environment: the hypothetical contexts of constructive dilemma.

Suppose you know that either John or Niko is in his office. You also know that if John is in his office then it must be Monday, and that if Niko is in his office then it must be Friday, but you do not know which day it is. After reflecting that either John or Niko is in his office, you suppose that John is in his office. Applying *modus ponens* in this hypothetical context, you infer that it must be Monday. You then suppose that Niko is in his office and infer that it must be Friday. By constructive dilemma, you conclude that either it must be Monday or it must be Friday. But given your evidence, it might not be Monday and it might not be Friday.

⁷There are other good forms of indirect proof besides. Here I present only the simplest ones where you suppose atomic A is true.

⁸Kolodny and MacFarlane also present “a more powerful variant” of McGee’s famous ‘counterexample’ to *modus ponens* (see Chapter 3, n. 4), but it involves the infelicitous use of *reductio* with a non-persistent premise of the form $\neg(\varphi \Rightarrow \psi)$. Again, *modus ponens* is not to blame.

Again, something has gone very wrong.⁹

| | | |
|---|---|----------------|
| 1 | John is in or Niko is in | Premise |
| 2 | If John is in, it must be Monday | Premise |
| 3 | If Niko is in, it must be Friday | Premise |
| 4 | <div style="border-top: 1px solid black; padding-top: 2px;">John is in</div> | Supposition |
| 5 | <div style="border-left: 1px solid black; padding-left: 5px;">It must be Monday</div> | From 2,4 |
| 6 | <div style="border-top: 1px solid black; padding-top: 2px;">Niko is in</div> | Supposition |
| 7 | <div style="border-left: 1px solid black; padding-left: 5px;">It must be Friday</div> | From 3,6 |
| 8 | It must be Monday or it must be Friday | From 1,4-5,6-7 |

It might be tempting here to think that the applications of *modus ponens* at step 5 and step 7 cause the trouble. But again, I think this misdiagnoses the problem, which arises only at the final step. And again, I think this misdiagnosis stems from a failure to appreciate the informational background of your deliberation. In the hypothetical context triggered by your supposition that John is in his office, your information leaves open the possibility only that John is in his office on a Monday, so is therefore information according to which either it must be Monday or it must be Friday. Likewise, in the hypothetical context triggered by your supposition that Niko is in his office, your information leaves open the possibility only that Niko is in his office on a Friday, so is therefore information according to which either it must be Monday or it must be Friday. But it does not follow from this that your original information is information according to which it must be Monday or it must be Friday. Information that leaves open the possibility only that either John or Niko is in his office needn't be information that either leaves open the possibility only that John is in his office or leaves open the possibility only that Niko is in his office.

Suppose instead that you know that either John must be in his office or Niko must be in his office. You also know that if John is in his office then it's Monday, and that if Niko is in his office then it's Friday. In this

⁹A similar example might lead one to worry that the argument from φ to $\Box\varphi$ —another form of argument that is validated by \models_I but not by \models_{Tr} —is unreliable in hypothetical contexts. Using this argument form and constructive dilemma, you can reason from the premise that either John or Niko is in his office to the conclusion that either John must be in his office or Niko must be in his office. However, I do not think that the argument from φ to $\Box\varphi$ is the problem. My diagnosis of this example and the one in the main text is the same.

case, you *can* felicitously conclude that either it must be Monday or it must be Friday:

| | | |
|---|--|----------------|
| 1 | John must be in or Niko must be in | Premise |
| 2 | If John is in, it's Monday | Premise |
| 3 | If Niko is in, it's Friday | Premise |
| 4 | John is in | Supposition |
| 5 | It's Monday | From 2,4 |
| 6 | Niko is in | Supposition |
| 7 | It's Friday | From 3,6 |
| 8 | It must be Monday or it must be Friday | From 1,4-5,6-7 |

If your information in the first hypothetical context has a structural feature—ruling out that it's not Monday—and your information in the second hypothetical context has another structural feature—ruling out that it's not Friday—then your initial information has at least one of these two features.

Importantly, there are also cases where you can felicitously use the premise that either John or Niko is in his office as input for constructive dilemma:

| | | |
|---|----------------------------|----------------|
| 1 | John is in or Niko is in | Premise |
| 2 | If John is in, it's Monday | Premise |
| 3 | If Niko is in, it's Friday | Premise |
| 4 | John is in | Supposition |
| 5 | It's Monday | From 2,4 |
| 6 | Niko is in | Supposition |
| 7 | It's Friday | From 3,6 |
| 8 | It's Monday or Friday | From 1,4-5,6-7 |

This non-modal variation of the first example is fine. Why? In this case, you conclude, not that either it must be Monday or it must be Friday, but instead that either it's Monday or Friday—and this weaker conclusion *is* established by your hypothetical reasoning. From the fact that your information in the first hypothetical context leaves open the

possibility only that it's Monday, and your information in the second hypothetical context leaves open the possibility only that it's Friday, it does not follow that your original information has one or the other of these structural features. It does follow that every possibility left open by your original information is a possibility where either it's Monday or Friday.¹⁰

The lesson from these examples, like before, is not that *modus ponens* for the indicative is unreliable, but instead that we should distinguish between these different forms of constructive dilemma:¹¹

1-constructive dilemma: If C follows from A and Γ , and D follows from B and Γ , then $\Box C \vee \Box D$ follows from $\Box A \vee \Box B$ and Γ .

2-constructive dilemma: If C follows from A and only *persistent* members of Γ , and D follows from B and only *persistent* members of Γ , then $C \vee D$ follows from $A \vee B$ and Γ .¹²

1-constructive dilemma and 2-constructive dilemma are used in the second and third example respectively. But the constructive dilemma in the first example fits neither of these forms. (The system **Info** in Appendix B is also sensitive to the distinction between 1-constructive dilemma and 2-constructive dilemma.)

5.2 Willer

So much for the first wave of attack. However, my defense of *modus ponens* is not yet complete. On the basis of Thomason conditionals and Moore's Paradox, Malte Willer [2010] also argues that *modus ponens* for the indicative conditional is invalid.¹³ I now argue that this second attack also fails. Willer misjudges the probative force of a premise in

¹⁰Let i^* explicate your information. It rules out the possibility that neither John nor Niko is in his office, rules out the possibility that John is in his office and it's not Monday, and rules out the possibility that Niko is in his office and it's not Friday, so $i^* \triangleright J \vee N$, $i^* \triangleright J \Rightarrow M$, and $i^* \triangleright N \Rightarrow F$. Consider some arbitrary $w^* \in i^*$. Since $i^* \triangleright J \vee N$, $\mathcal{V}(J, w^*) = T$ or $\mathcal{V}(N, w^*) = T$. Since $i^* \triangleright J \Rightarrow M$, $\mathcal{V}(J, w^*) = T$ only if $\mathcal{V}(M, w^*) = T$. Since $i^* \triangleright N \Rightarrow F$, $\mathcal{V}(N, w^*) = T$ only if $\mathcal{V}(F, w^*) = T$. Hence either $\mathcal{V}(M, w^*) = T$ or $\mathcal{V}(F, w^*) = T$, and since w^* was arbitrary, $i^* \triangleright M \vee F$.

¹¹I present only the simplest forms of constructive dilemma where you infer C from A and infer D from B .

¹²Without the restriction to persistent members, 2-constructive dilemma would lead to bad results. From the premise that either John or Niko is in his office, and the premise that John might not be in his office, you could conclude that Niko is in his office.

¹³Thomason's original example appears in van Fraassen [1980].

his alleged ‘counterexample’ to *modus ponens*. Willer’s argument, as follows, does not in fact show that this rule of inference is unreliable.

5.2.1 Attack

Suppose you find Sally rather cunning. In particular, you accept this Thomason conditional:

- (1) If Sally is lying, then I don’t believe that she’s lying.¹⁴

Further, suppose that you are a rational reflective agent. Not only do you fail to accept the Moore-paradoxical conjunction ‘Sally is lying and I don’t believe it,’ in fact you accept its negation:

- (2) It’s not the case that both Sally is lying and I don’t believe it.

You then reason as follows:

| | | |
|---|--|-------------|
| 1 | If Sally is lying, then I don’t believe it | Premise |
| 2 | Not: Sally is lying and I don’t believe it | Premise |
| 3 | Sally is lying | Supposition |
| 4 | I don’t believe that Sally is lying | From 1,3 |
| 5 | Sally is lying and I don’t believe it | From 3,4 |
| 6 | \perp | From 2,5 |
| 7 | Sally isn’t lying | From 3-6 |

In the hypothetical context triggered by your supposition that Sally is lying, you infer by *modus ponens* that you don’t believe that she is lying, and therefore that Sally is lying and you don’t believe it. Recognizing that this conflicts with the second premise, you conclude by *reductio* that Sally isn’t lying.¹⁵

But this inference, Willer claims, is infelicitous:

¹⁴Willer himself employs the indicative conditional ‘If Sally is deceiving me, then I don’t believe it.’ However, ‘deceive’ is arguably an *achievement* or *success* verb—the perlocutionary effect of deceit is the induction of inaccurate doxastic attitudes into the duped party—so Willer’s conditional is arguably an analytic truth. To avoid distracting complications stemming from this, I’ve substituted the non-success verb ‘lie.’

¹⁵Willer has you infer directly from the first premise that either Sally isn’t lying or you don’t believe it (on Willer’s formulation of *modus ponens*, this rule licenses a transition from an indicative conditional to the corresponding material conditional), then infer directly from the second premise that either Sally isn’t lying or you believe it, and then finally conclude that Sally isn’t lying. I suspect that Willer avoids my

This is the wrong result. Certainly, clever women are not always loyal, so [you] should not be allowed to infer Sally’s loyalty from her cleverness. [You have] very good reason to believe (1) and (2)...but [*modus ponens*] leads to a conclusion [you do] not have very good reason to believe. (p. 298)

Willer concludes that *modus ponens* is logically invalid. Importantly, he does *not* conclude that this argument form is invalid because it fails to unrestrictedly preserve truth—Willer is sympathetic to the idea that declarative sentences involving the indicative conditional are not truth-value bearers, and so semantic laws concerned with truth preservation do not apply to these sentences.¹⁶ He concludes rather that *modus ponens* is invalid because it licenses an inference from premises that you have good reason to believe to a conclusion that you do not have good reason to believe.

5.2.2 Defense

Now, I don’t think that this attack succeeds. To see why, let me apply the informational framework introduced in previous sections to Willer’s argument. One can, I think, see the error of Willer’s ways outside this formal framework—his mistake is basically just to underestimate the gulf between not believing and belief in the negation. But the formalism will help sharpen my diagnosis. While my primary aim in this subsection is to defend *modus ponens* against Willer’s attack, a secondary aim is to also show how the informational view of logic and deductive inquiry is useful for assessing deductive argumentation involving belief reports.

Let the sentence letter H abbreviate ‘Sally is lying to Harry’¹⁷ and let us add to \mathcal{L} a belief operator Bel used to designate what Harry believes. A model $\mathcal{M} = \langle \mathcal{W}, \mathcal{B}, \mathcal{V} \rangle$ for \mathcal{L} now consists of a set of possible worlds \mathcal{W} , an interpretation function $\mathcal{V} : At_{\mathcal{L}} \times \mathcal{W} \mapsto \{T, F\}$, and another function $\mathcal{B} : \mathcal{W} \mapsto 2^{\mathcal{W}}$ mapping each world to an information state.¹⁸ To determine $\llbracket \cdot \rrbracket_{\mathcal{M}} : \mathcal{S}_{\mathcal{L}} \times \mathcal{W} \times 2^{\mathcal{W}} \mapsto \{T, F\}$, the recursive specification of

reductio formulation to avoid the objection that in the hypothetical context where you suppose that Sally is lying you should no longer accept the first premise. However, this isn’t *my* objection and I find the *reductio* formulation more elegant.

¹⁶I’m also sympathetic to this idea. The case for expressivism in Chapter 3 focused on informational modal operators, but much the same goes for indicative conditionals.

¹⁷I’ve replaced the first-person pronoun ‘me’ with the proper name ‘Harry’ to sidestep distracting indexicality issues.

¹⁸Alternatively, one might cling to the standard relational semantics for modal languages (see Chapter 2, n. 9) by first introducing a binary accessibility relation $\mathcal{R} \subseteq \mathcal{W} \times \mathcal{W}$ between worlds and then letting $\mathcal{B}(w) = \{v : w\mathcal{R}v\}$.

truth in Chapter 2 is supplemented with this clause for Bel (cf. Yalcin [2011]):

$$\llbracket Bel(\varphi) \rrbracket_{\mathcal{M}}^{w,i} = T \quad \text{iff} \quad \mathcal{B}(w) \triangleright \varphi$$

The belief report ‘Harry believes that Sally is lying to him’ is true at $\langle w, i \rangle$ just in case the set $\mathcal{B}(w) \subseteq \mathcal{W}$ supplied by \mathcal{M} that explicates Harry’s belief state in w incorporates the embedded sentence H —that is, just in case Sally is lying to Harry in all of the ways the world might be that are left open by what Harry believes in w .¹⁹

With this semantics under our belts, we are well-positioned to see where Willer’s argument goes wrong.

Recall the first assumption that Harry believes that if Sally is lying to him then he doesn’t believe it—formally, $\mathcal{B}(@) \triangleright H \Rightarrow \neg Bel(H)$ where @ designates the actual world.

The second assumption is that Harry is rational and reflective.²⁰ In our possible worlds framework, Harry’s belief state $\mathcal{B}(w)$ in each world $w \in \mathcal{W}$ is nonempty and has the following *reflectivity* property:

Def 14. Belief state $\mathcal{B}(w)$ is reflective iff $\forall w' \in \mathcal{B}(w)(\mathcal{B}(w') = \mathcal{B}(w))$.²¹

If $\mathcal{B}(w)$ is reflective, then it is easy to check that for all $\varphi \in S_{\mathcal{L}}$, $\mathcal{B}(w) \triangleright \varphi$ iff $\mathcal{B}(w) \triangleright Bel(\varphi)$.²² Moreover, if $\mathcal{B}(w)$ is both reflective and nonempty,

¹⁹Modeling Harry’s belief state as a single set of possible worlds is admittedly unrealistic. Given the semantic clause for belief reports, either Harry’s beliefs are logically consistent and closed under logical consequence (when $\mathcal{B}(w)$ is nonempty) or he believes everything (when $\mathcal{B}(w) = \emptyset$). On a more realistic picture allowing for logical incoherence, Harry’s doxastic state is *fragmented*. His total doxastic state might be modeled as a *subset* of $2^{\mathcal{W}}$ (Stalnaker [1984]) or as a function from partitions to subpartitions of \mathcal{W} (Yalcin [2011]). For present purposes, though, this more complex structure is unnecessary; the simplest, most idealized model will do.

²⁰According to Willer, reflective agents have perfect higher-order knowledge of what they believe and do not believe: “We want agents to be reflective, their belief sets encoding not only first-order beliefs but also being closed under what the agent considers to be his own doxastic state” (p. 295). Willer himself cashes this out in a syntactic framework where belief states are sets of sentences but, as I’ll now show, reflectivity can also be modeled in a possible worlds framework.

²¹‘Reflective’ as used in Def 14 is a technical term applicable to the mathematical objects explicating what Harry believes, not an informal term applicable to Harry himself. But the formal and informal uses are related: Harry is reflective in w iff $\mathcal{B}(w)$ is reflective.

On the standard relational semantics (see n. 18), the reflectivity requirement amounts to the requirement that \mathcal{R} is *transitive* ($\forall w, v, u((w\mathcal{R}v \wedge v\mathcal{R}u) \supset w\mathcal{R}u)$) and *Euclidean* ($\forall w, v, u((w\mathcal{R}v \wedge w\mathcal{R}u) \supset v\mathcal{R}u)$). The requirement that $\mathcal{B}(w)$ is nonempty amounts to the requirement that \mathcal{R} is *serial* ($\forall w \exists v(w\mathcal{R}v)$).

²²This equivalence holds when $\mathcal{B}(w) = \emptyset$ since \emptyset incorporates all $\varphi \in S_{\mathcal{L}}$. In general, $\mathcal{B}(w) \triangleright \varphi$ iff $\forall w' \in \mathcal{B}(w)(\mathcal{B}(w') \triangleright \varphi)$ iff $\forall w' \in \mathcal{B}(w)(\llbracket Bel(\varphi) \rrbracket_{\mathcal{M}}^{w',\mathcal{B}(w)} = T)$

then for all $\varphi \in S_{\mathcal{L}}$, $\mathcal{B}(w) \not\vdash \varphi$ iff $\mathcal{B}(w) \triangleright \neg Bel(\varphi)$.²³ So if Harry believes such and such then he believes that he believes it, and if Harry does not believe such and such then he believes that he does not believe it.

Because Harry is a rational reflective agent, Willer thinks that Harry must also believe that it's not the case that Sally is lying and he doesn't believe it—formally, $\mathcal{B}(@) \triangleright \neg(H \wedge \neg Bel(H))$:

Moore paradoxical constructions are unacceptable to agents who are rational. Unacceptability comes in different flavors. The agent might be unable to accept φ since he lacks sufficient evidence in support of φ . The agent might also be unable to accept φ since φ is a priori absurd. For instance, φ might be an obvious contradiction. In such cases, not only is φ unacceptable, but also is the agent rationally committed to accept $\neg\varphi$. Neither accepting φ nor $\neg\varphi$ is not an option: Unacceptability of φ commits any rational agent to acceptance of $\neg\varphi$. Moorean paradoxical constructions are unacceptable in the latter sense: It is simply absurd for a rational agent to judge true both that φ holds and that he does not believe that φ holds. (p. 298)

But this is, I submit, Willer's error. I'll happily grant that a rational reflective agent cannot accept Moore paradoxical constructions. Since $\mathcal{B}(@)$ is nonempty and reflective, for instance, $\mathcal{B}(@) \not\vdash H \wedge \neg Bel(H)$.²⁴ However, I will not grant that a rational reflective agent must accept the negations of these paradoxical constructions. Information states that do not incorporate φ will incorporate $\diamond\neg\varphi$, but $\triangleright\neg\varphi$ is typically a far stronger constraint than $\triangleright\diamond\neg\varphi$. If $\mathcal{B}(@) \triangleright \neg(H \wedge \neg Bel(H))$ and $\mathcal{B}(@)$ is reflective, either $\mathcal{B}(@) \triangleright \neg H$ or $\mathcal{B}(@) \triangleright H$ —that is, either Harry's belief state rules out the possibility that Sally is lying, or it rules out the possibility that Sally isn't lying. Surely rationality and reflectivity alone cannot mandate that Harry's belief state has this structure.

Indeed, if Harry's belief state *has* this structure, then he *does well* to infer that Sally isn't lying to him. Putting the initial two assumptions together, his belief state cannot rule out the possibility that Sally isn't lying: if $\mathcal{B}(@) \triangleright H$, then $\mathcal{B}(@) + H = \mathcal{B}(@)$ and $\mathcal{B}(@) \triangleright Bel(H)$ by

iff $\mathcal{B}(w) \triangleright Bel(\varphi)$.

²³This equivalence fails when $\mathcal{B}(w) = \emptyset$ since $\emptyset \triangleright \varphi \wedge \neg Bel(\varphi)$. When $\mathcal{B}(w) \neq \emptyset$, $\mathcal{B}(w) \not\vdash \varphi$ iff $\forall w' \in \mathcal{B}(w) (\mathcal{B}(w') \not\vdash \varphi)$ iff $\forall w' \in \mathcal{B}(w) (\llbracket Bel(\varphi) \rrbracket_{\mathcal{M}}^{w', \mathcal{B}(w)} = F)$ iff $\forall w' \in \mathcal{B}(w) (\llbracket \neg Bel(\varphi) \rrbracket_{\mathcal{M}}^{w', \mathcal{B}(w)} = T)$ iff $\mathcal{B}(w) \triangleright \neg Bel(\varphi)$.

²⁴Suppose $\mathcal{B}(@) \triangleright H \wedge \neg Bel(H)$. Since $\mathcal{B}(@) \triangleright H$ and $\mathcal{B}(@)$ is also reflective, $\mathcal{B}(@) \triangleright Bel(H)$. But then $\mathcal{B}(@) \triangleright Bel(H) \wedge \neg Bel(H)$, contradicting that $\mathcal{B}(@) \neq \emptyset$. Similar reasoning establishes that $\mathcal{B}(@) \not\vdash Bel(H) \wedge \neg H$, $\mathcal{B}(@) \not\vdash H \wedge \diamond\neg Bel(H)$, $\mathcal{B}(@) \not\vdash H \wedge Bel(\diamond\neg H)$, and so forth.

reflectivity, so contrary to the first assumption $\mathcal{B}(@) \not\triangleright H \Rightarrow \neg Bel(H)$. Since $\mathcal{B}(@) \not\triangleright H$, $\mathcal{B}(@) \triangleright \neg H$. A nonempty reflective belief state that incorporates the two premises of Willer’s argument is therefore also an information state that rules out the possibility that Sally is lying to Harry. To say, as Willer does, that Harry mistakenly infers Sally’s loyalty from her cleverness is to ignore the crucial role of the second premise in the argument.

Even if $\mathcal{B}(@)$ is not reflective, Harry does well to argue from the premises $H \Rightarrow \neg Bel(H)$ and $\neg(H \wedge \neg Bel(H))$ to the conclusion $\neg H$. If $\mathcal{B}(@) \triangleright H \Rightarrow \neg Bel(H)$, then Harry’s belief state rules out the possibility that Sally is lying and he believes it. If $\mathcal{B}(@) \triangleright \neg(H \wedge \neg Bel(H))$, then his belief state also rules out the possibility that Sally is lying and he doesn’t believe it. Consequently, *any* belief state that incorporates both premises of the argument must incorporate the conclusion that Sally isn’t lying to Harry.

Willer’s argument, I conclude, does not give us good reason to think that *modus ponens* for the indicative is unreliable. The application of this rule does not lead Harry to an unwarranted conclusion. In fact, I think we can say a bit more: not only is the argument from $H \Rightarrow \neg Bel(H)$ and $\neg(H \wedge \neg Bel(H))$ to $\neg H$ a *good argument*, but the *reductio* in §5.2.1 is *good argumentation*. Consider the informational background of this episode. When Harry supposes that Sally is lying, he thereby enters a hypothetical context where the open possibilities are all ones in which Sally is lying. Since Harry’s information in this context still rules out the possibility that Sally is lying and he believes it, he does well to recognize that these live possibilities are also ones where he doesn’t believe that Sally is lying. However, Harry’s information also rules out the possibility that Sally is lying and he doesn’t believe it. There are in point of fact *no* open possibilities where Sally is lying—Harry’s degenerate information in the hypothetical context rules out everything. Thus, he does well to conclude that Sally is loyal.²⁵ Far from leading Harry astray, *modus ponens* helps establish what is so according to what he believes in @.²⁶

²⁵In §5.1.1, I warned that one must be careful when using *reductio ad absurdum* in languages that include informational constants. If you suppose H and derive \perp , then you can safely conclude $\Diamond \neg H$. However, you can conclude $\neg H$ only if every sentence φ imported into the subproof is *persistent*. In the present case, Harry can draw this stronger conclusion since both $H \Rightarrow \neg Bel(H)$ and $\neg(H \wedge \neg Bel(H))$ are persistent.

²⁶On the simple model I’ve been working with, Harry *already* believes that Sally is loyal before he engages in deductive inquiry. To make sense of how *reductio* can be a learning experience in which Harry acquires new beliefs, we might model his belief state $\mathcal{B}(@)$ as a set of fragments $\{\mathcal{B}(@)_n\}_{n \leq N}$ (recall n. 19) and upgrade the semantic clause for belief reports accordingly: $\llbracket Bel(\varphi) \rrbracket_{\mathcal{M}}^{w,i} = T$ iff $\exists n(\mathcal{B}(w)_n \triangleright \varphi)$.

Prior to Harry's deliberation, the story might now go, $\mathcal{B}(@) = \{\mathcal{B}(@)_1, \mathcal{B}(@)_2\}$ where $\mathcal{B}(@)_1 \triangleright H \Rightarrow \neg Bel(H)$ and $\mathcal{B}(@)_2 \triangleright \neg(H \wedge \neg Bel(H))$, but $\mathcal{B}(@)_1 \not\triangleright \neg H$ and $\mathcal{B}(@)_2 \not\triangleright \neg H$. The indirect proof removes this fragmentation. After deliberation, $\mathcal{B}(@) = \{\mathcal{B}(@)_3\}$ where $\mathcal{B}(@)_3 \triangleright H \Rightarrow \neg Bel(H)$, $\mathcal{B}(@)_3 \triangleright \neg(H \wedge \neg Bel(H))$, and $\mathcal{B}(@)_3 \triangleright \neg H$.

Chapter 6

The Myth of Logical Coherence

I've claimed in preceding chapters that logical validity, understood along informational lines, coincides with good deductive argument. But my case so far has been piecemeal. In Chapter 3, I claimed that the formal consequence relation \models_I validates a handful of good arguments involving informational constants that are invalidated by \models_{Tr} . In Chapter 5, I defended the reliability of *modus ponens* for the indicative conditional in hypothetical contexts. In this later chapter, though, I also discussed the *informational background* of theoretical deliberation, and so gestured at a general explanation for the equivalence of Valid_I and Valid_G .

Deductive argumentation, on the informal, pre-theoretic picture I have had in mind, is an information-driven enterprise in which an agent investigates what is so according to a salient body of information that incorporates the premises of an argument. In many contexts, this body of information is the informational content of the agent's beliefs—the agent is trying to determine how things actually are in the world. But it needn't be. An agent might be investigating what is so according to the clues in the famous zebra puzzle, or a politician's stump speech, or the testimony of an untrustworthy eyewitness to the murder, and so forth. If this testimony incorporates that Colonel Mustard did it and that if Colonel Mustard did it then he used the revolver, what else does this information incorporate?

One answer, of course, is that Colonel Mustard used the revolver. Another is that either Colonel Mustard used the revolver or Reverend Green used the candlestick. In general, the inferences that the agent does well to make on the basis of this testimony are precisely those that preserve incorporation. My punch line is already, I hope, obvious: on the natural, intuitive picture of deductive argumentation just sketched, the *good* deductive arguments are precisely those that count as valid on the informational view. The fact that $\{\neg A\} \models_I \neg \Diamond A$, $\{A \vee B, \neg B\} \models_I \Box A$, and $\{A \Rightarrow (\neg B \Rightarrow C), A\} \models_I \neg B \Rightarrow C$ is not a happy coincidence.

A clear picture of what is going on in deductive argumentation reveals that the informational concept of logical validity coincides with good deductive argument.

But what about the third normative conception of validity from Chapter 1?

Valid_N The argument from $\varphi_1, \dots, \varphi_n$ to ψ is logically valid if and only if it plays a special normative role in our epistemic and/or linguistic practices—for instance, the argument is valid just in case, subject to meeting certain qualifications, it determines a normative constraint of deductive cogency on thought *as such*.

We should distinguish these two questions:

- What is the connection between logical validity and good deductive argument?
- What is the normative import, if any, of logical validity within the first-person standpoint of deliberation and the second-person standpoint of advice?

While I have said much in response to the first question, I have said very little in response to the second question. So let us now inquire into the normativity of logic for thought and related intentional activity. Do the conceptions Valid_I and Valid_N also line up? By adopting the informational view of logic, can we hold onto a quasi-Fregean picture on which logic is a descriptive science that tells us which deductive arguments are good and also informs, in some special sense, what we ought to do and believe?

The normativity of logic is a difficult, thorny issue. I will not do it full justice here. But I want to critically examine one widely accepted proposal at the logic-epistemology interface—namely, that there exist deliberative requirements to have full beliefs that are logically consistent and closed under logical consequence, whatever else might be the case. I argue that this proposal does not hold up under scrutiny.¹

6.1 Bridging Logic and Epistemology

Lewis Carroll's [1895] Achilles is trying to get the Tortoise to accept the conclusion of some *modus ponens* arguments. But the Tortoise, ever the

¹I should say that my discussion in this chapter is largely independent of the informational view of logic. My main points should still be of interest to readers who cling to some version of the truth preservation view.

wisecrack, is being obstinate. He repeatedly accepts the premises of each argument but, to Achilles' increasing frustration, he refuses to accept the conclusion. At one point in their dialogue, Achilles finally loses his cool and says something rather curious:

Then Logic would take you by the throat and force you to do it!...Logic would tell you 'You can't help yourself. Now that you've accepted [the premises], you must accept [the conclusion]!' So you've no choice, you see. (p. 280)

Now, Achilles mistakenly conflates logic and epistemology. Logic, by itself, tells us nothing about what we are required, permitted, and forbidden to believe.² As Gilbert Harman puts the point in *Change in View* [1986], logic is not a theory of reasoning, at least not if reasoning is broadly conceived as a non-monotonic organic process of forming, maintaining, and revising beliefs over time—what he calls “reasoned change in view”—and not just as the churning out of consequences from a given set of premises—what he calls “argument” or “proof.” There is a *gap* between logic and reasoned change in view. Bridge principles are required that link up logical relations with requirements on full belief.

Still, is Achilles' remark, though unduly dramatic, completely off the mark? Isn't the Tortoise violating some requirement by accepting the premises of each argument but refusing to accept the conclusion? Wouldn't the Tortoise come to satisfy some requirement by accepting the conclusion or revising his belief in at least one of the premises? Harman argues that various candidate bridge principles linking logic and reasoning are all problematic, so he negatively concludes that logic plays no special role in reasoned change in view. Post *Change in View*, however, some philosophers—for example, Broome [1999], MacFarlane [*ms.*], Field [2009a] and [*ms.*]—have thought otherwise and proposed tight connections between logical relations and theoretical reasoning in the broad sense.

I focus here on what Kolodny [2007] calls “rational requirements of formal coherence *as such*” on full belief. Here are the two simplest requirements that bind at each moment in time:

Non-Contradiction: Rationality requires you either not to believe that φ is true, or not to believe that $\neg\varphi$ is true.³

²Recall from §4.1 that Hartry Field disagrees. But I'm now understanding logic along informational lines.

³Formally: $O_R(\neg Bel(\varphi) \vee \neg Bel(\neg\varphi))$ where O_R is the *ought of rationality* and $Bel(\varphi)$ designates that you believe that φ is true. Note that I'm using 'is true' here as a mere disquotational device.

Single-Premise Closure: Where φ logically implies ψ , rationality requires you either not to believe that φ is true, or to believe that ψ is true.⁴

These requirements have two key features. First, the rational ‘ought’ takes wide scope over the disjunctions. The principles do not require you to have or lack any one doxastic attitude but instead require that your beliefs logically cohere together in a particular way (exception: a closure principle can require belief in some logical truths). Second, these requirements are non-derivative principles that impose coherence constraints whatever else might be the case.⁵

My question is this: Are requirements like Non-Contradiction and Single-Premise Closure truly *deliberative requirements*? Does the fact that rationality requires you not to have contradictory beliefs, or to have beliefs that are closed under single-premise logical entailments constitute a reason, or entail that there is a reason, not to have logically incoherent combinations of attitudes?⁶

The inspiration for this chapter is a series of essays by Kolodny [2005], [2007] and [2008] in which he explores the conjecture that the scope of rational obligation is much narrower than is commonly thought. In his [2007], in particular, Kolodny tries to dispel “The Myth of Formal Coherence” by first pointing out just how puzzling various coherence

⁴Formally: Where $\{\varphi\} \models_I \psi$, $\text{OR}(\neg \text{Bel}(\varphi) \vee \text{Bel}(\psi))$. Though I take the relation of \models_I to be a set of sentences in $S_{\mathcal{L}}$ and a sentence in $S_{\mathcal{L}}$, I find it convenient to also talk of the logical consistency and closure of *belief*:

S ’s full beliefs are logically consistent iff where $\{\varphi_1, \dots, \varphi_n\} \models_I \perp$, either S does not believe that φ_1 is true ... or S does not believe that φ_n is true.

S ’s full beliefs are logically closed iff where $\{\varphi_1, \dots, \varphi_n\} \models_I \psi$, either S does not believe that φ_1 is true ... or S does not believe that φ_n is true, or S believes that ψ is true.

⁵The ‘*as such*’ in Kolodny’s characterization is crucial. It might be that you ought either not to believe that φ is true, or not to believe that $\neg\varphi$ is true simply because, given the evidence, you ought not to believe that $\neg\varphi$ is true. The force of the wide scope coherence norm is then parasitic on the force of the narrow scope evidential norm not to believe that $\neg\varphi$ is true which does not, by itself, require consistency *as such*. But rational requirements of coherence *as such* are not these derivative principles. Bracketing off all other requirements on belief, Non-Contradiction and Single-Premise Closure still require that your full beliefs are logically consistent and closed under single-premise logical entailments.

⁶Broome, MacFarlane, and Field all propose closure requirements with a wide scope ‘ought’ operator. But they are primarily concerned with the *content* of bridge principles and less so with their *force*. So it is not entirely clear that they all think of their proposed principles as rational requirements of formal coherence *as such*, let alone as *normative* requirements applicable in the first-person and second-person deliberative standpoints.

requirements are from the deliberative standpoint and then providing a *theory of error* for their existence. Roughly, the strategy behind the error theory is to explain our intuitions that particular formal coherence requirements exist by appealing to combinations of other requirements that do not individually require coherence *as such* yet together require that an agent's attitudes are coherent, or near coherent, in relevant ways. Since the force of rational requirements of formal coherence *as such* is not derivative on the force of any other requirements, explaining these coherence intuitions in terms of narrow scope principles that do not individually require coherence *as such* threatens the idea that the attitudes of a rational agent ought to be coherent in this or that fashion, whatever else might be the case.⁷

This chapter is longer than the others, so let me state the agenda. I begin in §6.2 by discussing a serious problem for Non-Contradiction and Single-Premise Closure, understood as deliberative requirements, that motivates the search for an error theory for their normativity. After then sketching Kolodny's theory in §6.3 and §6.4, I raise some trouble for it in §6.5. My main concern with his theory is not that it is incorrect, but rather that it is incomplete, unable to fully explain our intuitions that both Non-Contradiction and Single-Premise Closure are genuine deliberative requirements. In fact, I argue that a simpler error theory can do better. In §6.6, I enumerate desiderata for a satisfactory error theory and briefly discuss some aspects of theoretical inquiry that I will draw on. The simple theory is then presented in §6.7 and §6.8. In §6.9, I consider whether this theory overcomes my objections to Kolodny's original theory and I respond to some anticipated concerns with my theory. I conclude in §6.10.

6.2 Problems for Logical Requirements

It is worth first mentioning some minor worries with Non-Contradiction and Single-Premise Closure just to put them aside.

First, the unqualified Single-Premise Closure is too demanding. If Non-Contradiction and Single-Premise Closure are normative for thought, I take it that these principles apply only to what is considered, or ought to be considered, in deliberation. Still, the closure requirement seems

⁷Kolodny conjectures that requirements of formal coherence *as such* do not exist, but his real target, like mine, is the idea that principles like Non-Contradiction and Single-Premise Closure are normative for thought. His error theory aims at the idea that these principles apply within the first-person standpoint of deliberation and the second-person standpoint of advice and not just within the third-person standpoint of appraisal.

excessive when the premise is, say, a conjunction of axioms of Peano Arithmetic and the conclusion is some complex unproven theorem of PA. Presumably the scope of a closure requirement must be restricted in some way. However, I do not want to get into the mess of competing proposals.⁸ So I will just work with the unqualified requirement and point out when the oversimplification is problematic.

Second, Non-Contradiction and Single-Premise Closure are both *state requirements*. They specify logically incoherent conflict states to be avoided or escaped at a particular time. But one might insist that we should focus instead on *process requirements* that specify how one ought to go about changing one’s view over time—especially since reasoners can transition from one logically coherent view to another by all sorts of deviant routes.⁹ However, process requirements are more complicated and, in any case, doubt cast on the normativity of Non-Contradiction and Single-Premise Closure for thought is doubt cast on the normativity of process requirements geared at avoiding logically incoherent conflict states. So I will just work with the synchronic principles.

With these worries out of the way, let me turn to a more serious problem with the proposal that Non-Contradiction and Single-Premise Closure are deliberative requirements—what Kolodny [2007] calls “The Problem of Normativity.” Simply put, it is unclear *why* we should comply with these requirements:

The Problem of Normativity. Bracketing off other requirements on full belief except for Non-Contradiction and Single-Premise Closure, what *reasons* do you have to comply with these rational requirements? If no considerations that matter within your deliberation count in favor of logical coherence *as such*, then the *ought of rationality* O_R appearing in Non-Contradiction and Single-Premise Closure is divorced from the normative or deliberative *ought of reasons* O_T which concerns what you ought to believe or choose in deliberation.¹⁰

⁸Broome’s [1999] requirement pertains only to “immediate” logical entailments, MacFarlane’s [*ms.*] requirement pertains only to inferences that are “apprehended” as instances of a logically valid schema, and Field’s [2009a] requirement pertains only to “obvious” logical entailments. See their papers for details.

⁹Moreover, Non-Contradiction and Single-Premise Closure seem superfluous if there are also process requirements of logical coherence *as such* since a reasoner who adheres to process requirements organized around avoiding logically incoherent conflict states will also end up satisfying these synchronic requirements.

¹⁰The distinction between the *ought of rationality* and the *ought of reasons* is somewhat artificial. In Kolodny [2005], O_R and O_T are differentiated by the type of relations in their scopes: O_R operates on relations between an agent’s propositional attitudes, whereas O_T operates on relations between an agent’s attitudes and facts. However, Kolodny no longer endorses this way of drawing the distinction. First, he

The Problem of Normativity arises within the first-person standpoint of theoretical deliberation where you ask ‘What ought I to believe?’ Kolodny’s challenge is to answer the *deliberative question* ‘Why ought I to satisfy Non-Contradiction and Single-Premise Closure, whatever else might be the case?’

The natural responses are unsatisfying. First option: I ought to satisfy Non-Contradiction and Single-Premise Closure since complying with these logical coherence principles will lead me to believe the true and not to believe the false in any particular case, or at least will lead me to have full beliefs that are well supported by the evidence in any particular case. However, I might, say, restore consistency to my beliefs by dropping an accurate belief that is favored by the evidence while retaining an inaccurate belief that is not.

Second option: I ought to be disposed to satisfy Non-Contradiction and Single-Premise Closure since exercising these dispositions over time will lead me towards the true and away from the false in the long run, or at least will lead me to believe what is well supported by the evidence in the long run. However, this suggestion would explain only why I ought to cultivate dispositions to make my beliefs logically coherent *as such*, not why I ought to make my beliefs logically coherent in any particular case.¹¹

Third option: I ought to satisfy Non-Contradiction and Single-Premise Closure since having a high degree of logical coherence among certain of my attitudes is *constitutive* of being a believer and I have good reason to maintain this valuable status.¹² However, the threat of ceasing to

now thinks that O_r need not operate on relations between an agent’s propositional attitudes and facts. Second, he now thinks that “subject-evidence” principles that require that an agent’s attitudes cohere with her own evidence count as rational requirements. If this evidence is a body of facts, then O_R can operate on relations between an agent’s attitudes and facts. I’m carving the joints in this way: O_r is the ‘ought’ that concerns what you ought to believe or choose given the considerations that matter within deliberation, whereas O_R is just the ‘ought’ that features in purported ‘rational requirements.’ Note that on this demarcation, O_r and O_R can overlap since philosophers often apply the term ‘rational requirement’ to genuine deliberative requirements.

¹¹Moreover, Kolodny [2008] argues that exercising dispositions to satisfy Non-Contradiction and Single-Premise Closure in conjunction with other dispositions need not have these nice effects, or the dispositions to logical coherence *as such* would be superfluous given the required complementary dispositions.

¹²Davidson [2004] suggests that to interpret an agent as having beliefs at all, we must interpret her as having full beliefs that are largely free of logical incoherence. But he would not accept the present suggestion that this consideration can matter in deliberation:

Agents can’t *decide* whether or not to accept the fundamental attributes of ratio-

be a believer is a consideration that rarely, if ever, matters within one's deliberation. Moreover, this suggestion, if correct, only explains why I ought to comply with Non-Contradiction and Single-Premise Closure for the most part; again, it does not explain why I ought to make my beliefs coherent in any case.

Fourth option: I ought to satisfy Non-Contradiction and Single-Premise Closure since belief is *constitutively* governed by these norms. A propositional attitude counts as a *belief*, in part, because the holder of this attitude is subject to these logical coherence requirements. However, this second constitutive claim is more obscure and requires considerable argument. Moreover, even if this claim is true, it remains obscure how the constitutive norms Non-Contradiction and Single-Premise Closure can carry any normative force in theoretical deliberation—that is, how they engender reasons for belief.¹³

The Problem of Normativity is not decisive. It remains an open question whether Kolodny's challenge can be met.¹⁴ But I agree with Kolodny that our inability to make sense of Non-Contradiction and Single-Premise Closure from the deliberative standpoint motivates the search for an error theory. Note that a successful error theory would put additional pressure on the wide scope logical coherence requirements. Inside deliberation, they would seem redundant.

nality. If they are in a position to decide anything, they have those attributes.
(p. 196-7)

¹³See Kolodny [2007] for further discussion. At this point, one might reject the idea that Non-Contradiction and Single-Premise Closure are normative and insist that they are only evaluative standards for belief-forming rules or processes. From the third-person standpoint of appraisal, one might claim, an agent *functions well* by revising a contradictory belief or by closing her full beliefs under single-premise logical entailments, just as a heart that pumps blood at a certain rate or a fuel gauge that accurately measures the amount of gasoline in the tank is functioning well. However, this problem now arises:

The Problem of Appraisal. In what sense exactly does an agent function well by exercising a disposition to logical coherence *as such*?

It is not clear that this *evaluative question* can be satisfactorily answered either. In any case, this is not the question I'm interested in. Deliberators who recognize that they have logically incoherent beliefs will typically feel a kind of *normative pressure* to revise their incoherent attitudes. So my question is whether Non-Contradiction and Single-Premise Closure are truly normative requirements.

¹⁴I'm skeptical that it can. But I'm a bit less skeptical that the Problem of Appraisal in n. 13 can be satisfactorily addressed. Removing logical incoherence from your full beliefs might indicate correct use of a short-cut cognitive heuristic and so merit some positive appraisal. However, this line requires some empirical support.

6.3 Kolodny’s Error Theory for Non-Contradiction

The goal of Kolodny’s error theory for Non-Contradiction is to explain these violation and satisfaction claims:

Violation_N: If you believe that φ is true and you believe that $\neg\varphi$ is true, then you violate some requirement.

Satisfaction_N: If you believe that φ is true, you believe that $\neg\varphi$ is true, and you revise either of these beliefs, then you thereby satisfy some requirement that you would not satisfy if you retained both of these beliefs.

Kolodny’s [2007] main idea is this:

The attitudes that reason requires, in any given situation, are formally coherent. Thus, if one has formally incoherent attitudes, it follows that one must be violating some requirement of reason. The problem is not, as the idea of requirements of formal coherence *as such* suggests, that incoherent attitudes are at odds with *each other*. It is instead that when attitudes are incoherent, it follows that one of these attitudes is at odds with the *reason* for it—as it would be even if it were not part of an incoherent set. (p. 231)

6.3.1 Explaining Violation_N

Explaining the violation claim is straightforward. Let $e(\varphi)$ designate the degree to which your total evidence $\mathcal{E} \in 2^{\mathcal{W}}$ supports that φ is true.¹⁵ Then assuming that—

Stronger Evidence: Reason requires you not to believe that φ is true if $e(\neg\varphi) \geq e(\varphi)$ ¹⁶

we have by Trichotomy—

R₁: Reason requires you not to believe that φ is true or requires you not to believe that $\neg\varphi$ is true.¹⁷

¹⁵Although I will not commit myself to a particular account of evidence and evidential support in this chapter, I assume throughout that the degree $e(\varphi)$ to which \mathcal{E} supports that φ is true is a real number.

¹⁶Formally: If $e(\neg\varphi) \geq e(\varphi)$, then $O_r(\neg Bel(\varphi))$. It is implicitly assumed that the value of avoiding an inaccurate belief exceeds the value of having an accurate belief. Hence $O_r(\neg Bel(\varphi))$ if your evidence bearing on whether φ is true is evenly balanced. A kind of Evidentialism is also assumed where only evidence can be a reason for belief. If there are *pragmatic* reasons to believe, then Stronger Evidence can fail.

¹⁷Formally: $O_r(\neg Bel(\varphi)) \vee O_r(\neg Bel(\neg\varphi))$. Foley’s [2009] “non-contradiction” is

Thus, if you have contradictory beliefs, then you violate a requirement of reason. Whereas Non-Contradiction is a wide scope requirement that can be satisfied by revising your belief that φ is true or by revising your belief that $\neg\varphi$ is true, R_1 is a disjunction of narrow scope requirements. If reason requires you not to believe that φ is true but permits you to believe that $\neg\varphi$ is true, then you move *against reason* by maintaining your belief that φ is true but dropping your belief that $\neg\varphi$ is true. So Satisfaction_N cannot also be explained in terms of R_1 .

6.3.2 Explaining Satisfaction_N

Explaining the satisfaction claim is trickier. Kolodny's strategy is to appeal to second-order requirements that you reflect on what reason requires and bridge principles connecting your beliefs about this with requirements of rationality.

If you satisfy either of—

Second-Order Requirement₁: If you believe that φ is true, believe that $\neg\varphi$ is true, and it matters sufficiently whether φ is true, then reason requires you either to believe that reason requires you not to believe that φ is true, or to believe that reason requires you not to believe that $\neg\varphi$ is true

Second-Order Requirement₂: If you believe that φ is true, believe that $\neg\varphi$ is true, and it matters sufficiently whether φ is true, then reason requires you to attempt to decide which of these beliefs reason permits together with the applicable rational requirement(s) in—

Believed Reason: If you believe that reason requires you to believe that φ is true, then rationality requires you to believe that φ is true. If you believe that reason requires you not to believe that φ is true, then rationality requires you not to believe that φ is true. If you are deliberating, in response to a live doubt, about whether reason permits you to believe that φ is true, then rationality requires you not to believe that φ is true¹⁸

similar: If it is rational for you to believe that φ is true, then it cannot be rational for you to believe that $\neg\varphi$ is true.

¹⁸For ease of exposition and since it makes little difference for present purposes, I deviate slightly from Kolodny's theory in presenting the rational requirements in Believed Reason as state requirements rather than process requirements. Formally: If $Bel(O_r(Bel(\varphi)))$, then $O_R(Bel(\varphi))$. If $Bel(O_r(\neg Bel(\varphi)))$, then $O_R(\neg Bel(\varphi))$. If you are deliberating, in response to a live doubt, about whether $O_r(\neg Bel(\varphi))$, then $O_R(\neg Bel(\varphi))$.

How should we understand the *ought of rationality* O_R in the consequents of the

then you will revise your belief that φ is true, revise your belief that $\neg\varphi$ is true, or both.¹⁹

Kolodny also takes the converse to hold: if you revise your belief that φ is true, revise your belief that $\neg\varphi$ is true, or both, then you satisfy at least one of the Second-Order Requirements and the applicable rational requirement(s) in Believed Reason.

6.4 Kolodny's Error Theory for Single-Premise Closure

The goal of the theory is to explain these violation and satisfaction claims:

Violation_{SPC}: Where φ logically implies ψ and you believe that φ is true but you do not believe that ψ is true, then you violate some requirement.

Satisfaction_{SPC}: Where φ logically implies ψ , you believe that φ is true, you do not believe that ψ is true, and you revise your belief that φ is true or form the belief that ψ is true, then you thereby satisfy some requirement that you would not satisfy if you left your beliefs alone.

6.4.1 Explaining Violations_{SPC}

Again, explaining the violation claim is straightforward. Assuming that—

conditionals in Believed Reason? If you believe inaccurately that reason requires you to believe that φ is true, what can be said for your coming to believe that φ is true? I discuss Kolodny's answer in §6.5.

¹⁹If you satisfy Second-Order Requirement₁, then you will believe that reason requires you to revise your belief that φ is true, or you will believe that reason requires you to revise your belief that $\neg\varphi$ is true. By the second conditional in Believed Reason, rationality requires you to revise your belief that φ is true or requires you to revise your belief that $\neg\varphi$ is true. If you satisfy Second-Order Requirement₂ and you attempt to decide which of your contradictory beliefs is permitted, then you will conclude, perhaps incorrectly, that reason requires you to revise your belief that φ is true, reason requires you to revise your belief that $\neg\varphi$ is true, reason requires you to revise both of these beliefs, or you will fail to reach a decision. By the second conditional in Believed Reason, rationality requires you to revise your belief that φ is true in the first case, revise your belief that $\neg\varphi$ is true in the second case, and revise both of your beliefs in the third case. By the third conditional in Believed Reason, rationality requires you to revise both of your beliefs in the fourth indecisive case.

Evidence Transmission: Where φ logically implies ψ , $e(\psi) \geq e(\varphi)$ ²⁰

Epistemic Strictness: Reason requires you to believe that φ is true or requires you not to believe that φ is true²¹

and the evidential demands for the question whether ψ is true are no higher than for the question whether φ is true²²—

R₂: Where φ logically implies ψ , reason requires you not to believe that φ is true or requires you to believe that ψ is true.²³

Thus, if your beliefs aren't closed across single-premise entailments, then you violate a requirement of reason.²⁴

6.4.2 Explaining Satisfaction_{SPC}

Kolodny does not actually explain Satisfaction_{SPC} but I presume that the explanation would go as follows.

If you satisfy either of—

Second-Order Requirement₃: Where φ logically implies ψ , you believe that φ is true, you do not believe that ψ is true, and it matters sufficiently whether φ is true and whether ψ is true, then reason requires you either to believe that reason requires you not to believe that φ is true, or to believe that reason requires you to believe that ψ is true

Second-Order Requirement₄: Where φ logically implies ψ , you believe that φ is true, you do not believe that ψ is true, and it matters sufficiently whether φ is true and whether ψ is true, then reason requires you to attempt to decide whether reason forbids believing that φ is true

²⁰As stated, ET is more of a closure principle than a transmission principle (in the sense of Wright [1985]). But I stick to Kolodny's terminology.

²¹Formally: $O_r(Bel(\varphi)) \vee O_r(\neg Bel(\varphi))$. Does ES require you to clutter your mind? Not according to Kolodny:

ES says that one is always required to make up one's mind in a particular way, if one makes it up. But one may be merely permitted to make up one's mind. (n. 39)

In other words, ES applies only when you consider whether φ is true in the course of your deliberation, but you need not consider this question when it is irrelevant or trivial. Note that ES, understood in this way, suffices for the error theory since I take it that Single-Premise Closure, understood as a deliberative requirement, only applies to what is considered, or ought to be considered, in deliberation.

²²More specifically, the threshold value of $e(\psi)$ above which reason permits you to believe that ψ is true is no higher than the corresponding threshold value for $e(\varphi)$.

²³Formally: Where $\{\varphi\} \models_I \psi$, $O_r(\neg Bel(\varphi)) \vee O_r(Bel(\psi))$.

²⁴Presumably R₂ must be qualified in some way—recall the mathematician who attends to a complex theorem of PA—but Violations_{SPC} must be similarly qualified.

and whether reason requires believing that ψ is true

together with the rational requirement(s) in Believed Reason, then you will revise your belief that φ is true or come to believe that ψ is true.

Kolodny also takes the converse to hold: if you revise your belief that φ is true or come to believe that ψ is true, then you satisfy at least one of the Second-Order Requirements and the applicable rational requirement(s) in Believed Reason.

Kolodny's conclusion is this:

How does logic govern belief? [A rational requirement of logical coherence *as such*] represents one answer: that logic somehow governs belief directly, such that if our beliefs are not consistent and closed, we violate some norm. Our discussion of R_1 and R_2 represents a different answer: that logic governs belief indirectly, by structuring epistemic reason, which in turn directly governs belief...The question, then, is whether, as [a rational requirement of logical coherence *as such*] implies, logic does double duty, not only structuring what epistemic reason requires, but also placing an independent constraint on belief that sometimes countermands what epistemic reason requires. This begins to seem like a fetish for a certain mental pattern. (p. 254-5)

6.5 Pressure Points

However, Kolodny does not claim to have dispelled The Myth of Formal Coherence:

I am not confident that I rule out the more pessimistic, if more interesting answer: that while we cannot find a place for requirements of formal coherence *as such*, we cannot do without them either. (p. 232)

The hedge is appropriate. Though the Problem of Normativity facing Non-Contradiction and Single-Premise Closure suggests that these are not truly deliberative requirements, there are many points of stress in Kolodny's error theory. Here are six of them:

No Required Self-Monitoring. Harman [1986] claims that in cases where you do not have much time to deal with a recognized logical inconsistency in your full beliefs and the cost of revision is high, your best response is to quarantine the inconsistent beliefs and move on. I

agree and worry that there will be many cases where Satisfaction_N or Satisfaction_{SPC} cannot be explained by any second-order requirements to self-monitor because, given strong pragmatic reasons to devote your scarce cognitive resources elsewhere, none of these requirements are in effect.

No Revision in Doubt. According to Kolodny [2005], the *ought of rationality* O_R in Believed Reason is not genuinely normative but is *transparent* to the *ought of reasons* O_r . From the outside, to tell someone who believes that reason requires her to believe that φ is true that rationality requires her to believe that φ is true is to make only the descriptive claim that, *from her point of view*, reason requires her to believe that φ is true. From the inside, this is experienced as sound advice because O_R looks just like O_r . The rational requirement will *seem* normative to the advisee since, from her perspective, what rationality and reason require coincide.

However, while Kolodny’s “Transparency Account” might explain the apparent normative force of the O_R appearing in the first and second conditionals of Believed Reason, this account cannot handle the third conditional. If you are deliberating, in response to a live doubt, about whether reason permits you to believe that φ is true, then you do *not* believe that reason requires you not to believe that φ is true. Rather, you have not decided whether reason requires this. So in this case at least, O_R is *not* transparent to O_r . The charge “Rationality forbids you to believe that φ is true!” cannot say, in effect, “As it seems to you, reason forbids you to believe that φ is true!” Indeed, a requirement not to believe that φ is true while in the process of deciding whether reason permits this belief seems like a substantive deliberative requirement—and recall Harman’s point that you needn’t revise a problematic belief if this would create a large disturbance in your overall web of beliefs. So I also worry about cases where Satisfaction_N or Satisfaction_{SPC} cannot be explained by the rational requirement in Believed Reason to revise in doubt.

Can’t Get No Satisfaction. Even in cases where you are required to self-monitor and revise in doubt, another serious worry remains. Assume that you believe that φ is true and you believe that $\neg\varphi$ is true. If you satisfy one of the second-order requirements and the applicable rational requirement(s) in Believed Reason, then you will revise your belief that φ is true, revise your belief that $\neg\varphi$ is true, or both. So far so good. But why should we accept the converse statement needed for the error theory to go through? If you revise your belief that φ is true, revise your belief that $\neg\varphi$ is true, or both, why does it follow that you satisfy one of the

second-order requirements and the applicable rational requirement(s) in Believed Reason? A crucial piece of the argument is missing.

General Case. To cover general logicity, Non-Contradiction and Single-Premise Closure must be supplemented with this principle:

Closure Under Conjunction: Rationality requires you either not to believe that φ is true, not to believe that ψ is true, or to believe that $\varphi \wedge \psi$ is true.²⁵

The three principles together imply the generalizations²⁶—

Consistency: Where $\varphi_1, \dots, \varphi_n$ are logically inconsistent, rationality requires you either not to believe that φ_1 is true, ..., or not to believe that φ_n is true.²⁷

Closure: Where $\varphi_1, \dots, \varphi_n$ logically imply ψ , rationality requires you either not to believe that φ_1 is true, ..., not to believe that φ_n is true, or to believe that ψ is true.²⁸

There are violation and satisfaction claims for Consistency and Closure akin to Violation_N, Satisfaction_{SPC}, and so on. Can a Kolodny-style error theory explain these claims?

Kolodny acknowledges that Preface and Lottery cases threaten the required generalizations of R₁ and R₂. On a common understanding of these paradoxical cases, your evidence \mathcal{E} can strongly support that each of the sentences $\varphi_1, \dots, \varphi_n$ appearing in the body of a text or expressing that lottery tickets will not win is true, but fail to support that their conjunction $\varphi_1 \wedge \dots \wedge \varphi_n$ is true—that is, $e(\varphi_1), \dots, e(\varphi_n)$ are all high but $e(\varphi_1 \wedge \dots \wedge \varphi_n)$ is low. Thus, reason might require you to believe that φ_1 is true, ..., require you to believe that φ_n is true, forbid you to believe that $\varphi_1 \wedge \dots \wedge \varphi_n$ is true, and even require you to believe that $\neg(\varphi_1 \wedge \dots \wedge \varphi_n)$ is true (cf. Christensen [2004] and Foley [2009]). However, Kolodny suggests that the violation claims in these cases are “overgeneralizations of a kind”: having developed logical coherence intuitions in garden-variety cases where disjunctions of narrow scope evidential norms are easily confused with wide scope requirements of logical coherence *as such*, we extend our intuitions to Preface and Lottery cases where reason does not in fact require us to have logically coherent full beliefs. The problem is not with his theory but with the violation claims themselves.

²⁵Formally: $O_R(\neg Bel(\varphi) \vee \neg Bel(\psi) \vee Bel(\varphi \wedge \psi))$.

²⁶Assuming that the inference from $O_R(\varphi \vee \psi)$ and $O_R(\neg\psi \vee \xi)$ to $O_R(\varphi \vee \xi)$ is valid. If you reject it, note that Consistency and Closure can be regarded as basic rather than derived principles.

²⁷Formally: Where $\{\varphi_1, \dots, \varphi_n\} \models_I \perp$, $O_R(\neg Bel(\varphi_1) \vee \dots \vee \neg Bel(\varphi_n))$.

²⁸Formally: Where $\{\varphi_1, \dots, \varphi_n\} \models_I \psi$, $O_R(\neg Bel(\varphi_1) \vee \dots \vee \neg Bel(\varphi_n) \vee Bel(\psi))$.

But what about non-paradoxical cases where the generalizations of R_1 and R_2 fail? For example, consider a two-premise case where you are considering whether (P1) Ellis Marsalis is performing at Carnegie Hall, whether (P2) if Ellis Marsalis is performing then the show is sold out, and whether (C) the show at Carnegie Hall is sold out. Suppose that the evidential demands for these questions are identical, $e(P1)$ and $e(P2)$ are just high enough so that reason permits you to believe that Ellis Marsalis is performing at Carnegie Hall and permits you to believe that if Ellis Marsalis is performing then the show is sold out, but $e(C) < e(P1)$ so reason forbids you to believe that the show at Carnegie Hall is sold out.²⁹ If these cases are common, then Kolodny must appeal to a high degree of overgeneralization to explain our violation intuitions.

Strictness. Epistemic Strictness is a very strong assumption. When considering whether φ is true, ES says that you must make up your mind in a particular way. However, can't reason simply permit but not require you to believe that φ is true in at least some cases? Suppose that reason requires you to believe that φ is true if only if $e(\varphi) \geq 0.8$. If ES holds and you are considering whether φ is true, then reason requires you not to believe that φ is true if $e(\varphi) < 0.8$. But can't there be a range of values for $e(\varphi)$ —say, $e(\varphi) \in (0.6, 0.8)$ —where reason does not require you to make up your mind?³⁰

Finally, this last worry is, to my mind, the most serious:

Pull of Coherence. Bracketing off my earlier worries, Kolodny's error theory still cannot explain the distinctive *feel* or *pull* of logical coherence. According to his theory, someone who believes that φ is true and believes that $\neg\varphi$ is true holds at least one of these beliefs against reason. The tension is between a contradictory belief and the evidence, not between the contradictory beliefs themselves. However, many of us have the stronger coherence intuition that this agent ought to revise her belief that φ is true *insofar as she continues to believe that $\neg\varphi$ is true* and vice versa. This coherence intuition is left unexplained.

²⁹Consider a toy model where $e : S_{\mathcal{L}} \mapsto [0, 1]$ is a probability measure, reason requires you to believe that φ is true if $e(\varphi) \geq 0.8$, and reason forbids you to believe that φ is true if $e(\varphi) < 0.8$. If $e(P1) = e(P2) = 0.8$, then reason requires you to believe that P1 is true and requires you to believe that P2 is true. However, unless $e(P1 \vee P2) = e(P1) = e(P2) = 0.8$, it is consistent with the Kolmogorov axioms that $e(C) < 0.8$ and reason forbids you to believe that C is true.

³⁰White [2005] argues for "Uniqueness": given your total evidence \mathcal{E} , there is a unique rational doxastic attitude that you must adopt towards each $\varphi \in S_{\mathcal{L}}$. But I will not take a firm stand in this debate. I'm content to make the weaker claim that ES is sufficiently controversial to warrant suspicion of theories that invoke it.

6.6 Towards Another Error Theory

Now, there are other pressure points in Kolodny's theory that I have not pressed, but I will stop here.³¹ A decision point has been reached: we can try harder to make sense of Non-Contradiction and Single-Premise Closure as deliberative requirements, or we can seek an error theory that avoids my worries in §6.5.

In these next few sections, I pursue the latter option. I am after an error theory that (I) does not rely unduly on second-order requirements to self-monitor, (II) does not rely on a rational requirement to revise in doubt, (III) can explain multi-premise violation and satisfaction claims, (IV) does not rely on Epistemic Strictness, and (V) can explain the distinctive pull of logical coherence. To construct such a theory, I'll draw on my informational account of theoretical inquiry and some general points about how reasoners form beliefs about what they ought and ought not to believe.

Suppose that you are investigating the murder of Mr. Boddy. Along with relevant background assumptions that you simply take for granted, you activate your belief that either Mrs. Peacock or Miss Scarlett did it. In doing so, you come to occupy a particular categorical context. From your categorical perspective, the world is such that, *inter alia*, Mrs. Peacock or Miss Scarlett is the murderess.³² Rightly or wrongly, you treat the information that either Mrs. Peacock or Miss Scarlett did it as evidence, including it in the informational basis that is driving your reasoning.³³

If you then consider whether to believe that Miss Scarlett is the murderess, the answer you give will depend on this basis. The question *whether to believe that φ is true* is *transparent* to the question *whether φ is true* in the sense that both questions, posed within deliberation, are answered by, and answerable to, the same set of considerations (cf. Shah [2003]). So your judgments about whether reason requires, permits, or forbids you to believe that Miss Scarlett is the murderess will depend, for the most part, on the extent to which you take the information behind your categorical perspective to support that Miss Scarlett is the

³¹For instance, one might worry about the Evidentialism behind SE (see n. 16).

³²In addition to categorical contexts, you can, of course, occupy hypothetical contexts in deliberation. You might have merely supposed that either Mrs. Peacock or Miss Scarlett did it. Whereas your goal in a categorical context is to figure out what is actually the case—you *identify* with your categorical perspective—your goal in a hypothetical context is to figure out what is or would be the case under certain suppositions.

³³In a possible worlds framework, this basis might be modeled by an information state $i \in 2^{\mathcal{W}}$.

murderess.

But one important caveat: though you regard this information as evidence plus background assumptions, presumably you will not come to believe that reason requires you to believe that Miss Scarlett is the murderess if you think the information in your categorical context omits relevant evidence bearing on whether she did it. After all, this ignored evidence might defeat or cancel the justificatory status of a belief that Miss Scarlett did it.

6.7 A Simple Error Theory for Non-Contradiction

On to my error theory. Recall the first pair of targets:

Violation_N: If you believe that φ is true and you believe that $\neg\varphi$ is true, then you violate some requirement.

Satisfaction_N: If you believe that φ is true, you believe that $\neg\varphi$ is true, and you revise either of these beliefs, then you thereby satisfy some requirement that you would not satisfy if you retained both of these beliefs.

A simpler explanation of these claims runs as follows. Suppose that you believe that φ is true, you believe that $\neg\varphi$ is true, and, without loss of generality, the former belief is active in your deliberation—that is, you are taking it to be part of your evidence that φ is true. From inside your categorical φ -perspective, you feel tension to revise your view upon recognizing that you also believe that $\neg\varphi$ is true. Again, the question *whether to believe that $\neg\varphi$ is true* is transparent to the question *whether $\neg\varphi$ is true* and, from the φ -perspective, the answer to this latter question is obvious. As you see the world, φ is true and $\neg\varphi$ is certainly false, so you believe that reason forbids you to believe that $\neg\varphi$ is true. Indeed, if you did not believe that reason forbids you to believe that $\neg\varphi$ is true, then presumably you could not even be said to actively believe that φ is true.

In believing that $\neg\varphi$ is true, you seemingly violate a narrow scope evidential norm. In revising this belief, you seemingly satisfy this norm. What if you revise your belief that φ is true instead? Well, if your other belief that $\neg\varphi$ is true is also active in deliberation, then you seemingly satisfy an evidential requirement not to believe that φ is true from inside the $\neg\varphi$ -perspective.³⁴ However, another explanation of the satisfaction

³⁴Though one might adopt both a categorical φ -perspective and a categorical $\neg\varphi$ -perspective in the course of one's deliberation, presumably one cannot adopt a categorical φ - $\neg\varphi$ -perspective. Inside any categorical perspective, a reasoner has a

intuition is readily available. By revising your belief that φ is true, you no longer identify with the φ -perspective, so it will no longer seem to you, on the basis of the truth of φ , that reason forbids you to believe that $\neg\varphi$ is true. Though you might not satisfy any believed requirement by removing the contradiction in your beliefs, you still seemingly *avoid violating* one. Given the ease with which satisfaction and non-violation intuitions are confused, I suggest that this is enough for the error theory for Non-Contradiction to go through.

6.8 A Simple Error Theory for Single-Premise Closure

Recall the second pair of targets:

Violation_{SPC}: Where φ logically implies ψ and you believe that φ is true but you do not believe that ψ is true, then you violate some requirement.

Satisfaction_{SPC}: Where φ logically implies ψ , you believe that φ is true, you do not believe that ψ is true, and you revise your belief that φ is true or form the belief that ψ is true, then you thereby satisfy some requirement that you would not satisfy if you left your beliefs alone.

The explanation of these claims is similar. Suppose that you believe that φ is true and this belief is active in your deliberation. When you consider whether ψ is true inside your categorical φ -perspective, you feel normative pressure to come to believe that ψ is true. From your perspective, ψ is certainly true, so you believe that reason requires you to believe that ψ is true.

In not believing that ψ is true, you seemingly violate a narrow scope evidential norm. In acquiring this belief, you seemingly satisfy this norm. What if you revise your belief that φ instead? Well, you no longer identify with the φ -perspective, so it will no longer seem to you, on the basis of the truth of φ , that reason requires you to believe that ψ is true. As it seems to you, you avoid violating an evidential norm. Moreover, in many cases where an agent believes that φ is true but does not believe that a considered ('obvious', 'immediate', etc.) logical entailment ψ is true, she also believes that $\neg\psi$ is true. From inside a $\neg\psi$ -perspective, it might also seem that you satisfy an evidential requirement by revising your belief that φ is true.

coherent, though possibly inaccurate, view of the world.

6.9 Virtues

That is my simple error theory for Non-Contradiction and Single-Premise Closure. By appealing only to what agents with logically incoherent full beliefs take their evidence to require, this theory avoids most, if not all, of my worries with Kolodny's original theory in §6.5.

First, my theory does not mention any second-order requirements to self-monitor, the rational requirement in Believed Reason to revise in doubt, and Epistemic Strictness.

Second, a natural extension of my theory can explain our general violation and satisfaction intuitions. For instance, recall the kind of multi-premise case in §6.5 where (I) φ_1 and φ_2 logically imply ψ , (II) you are considering whether φ_1 is true, whether φ_2 is true, and whether ψ is true, (III) the evidential demands for these questions are identical, (IV) $e(\varphi_1)$ and $e(\varphi_2)$ are just high enough so that reason permits you to believe that φ_1 is true and permits you to believe that φ_2 is true, (V) $e(\psi) < e(\varphi_1)$ so reason forbids you to believe that ψ is true. Here, reason permits you to have logically incoherent beliefs. Still, if you appreciate that your beliefs are incoherent from inside the φ_1 - φ_2 -perspective where you take it to be part of your evidence that $\varphi_1 \wedge \varphi_2$ is true, you will feel normative pressure to revise your view since it is certainly the case that ψ is true. Given your actual evidence \mathcal{E} , $e(\psi)$ is sufficiently low that reason forbids you to believe that ψ is true. However, as it seems to you within the φ_1 - φ_2 -perspective, reason requires you to believe that ψ is true, so the error theory goes through as before.

Third, my theory suggests an explanation of the distinctive pull of logical coherence. Suppose that you believe that φ is true, you believe that $\neg\varphi$ is true, and you recognize the contradiction in your beliefs from inside the $\neg\varphi$ -perspective where you believe that reason forbids you to believe that φ is true. Inside this categorical perspective, you might reflect: *given that $\neg\varphi$ is true*, I ought not to believe that φ is true. But this is not quite what we want. The tension is between your belief that φ is true and the apparent fact that $\neg\varphi$ is true, *not* between your contradictory beliefs themselves. But a combination of two slight confusions can get us what we want. The first confusion involves a subtle change in adverbial phrase. I suggest that you might think instead: *given that I believe that $\neg\varphi$ is true*, I ought not to believe that φ is true. As it seems to you, you satisfy a norm by revising your belief that φ is true, and you avoid violating a norm by revising your belief that $\neg\varphi$ is true. The second confusion is to mistake a non-violation claim for a satisfaction claim. I suggest that you might also think: *given that I believe that φ is true*, I ought not to believe that $\neg\varphi$ is true. The overall

result of these reflections is a pull towards logical consistency that can generate the idea that Non-Contradiction governs thought.

Similarly, if you recognize that your beliefs are not closed under logical consequence from inside the φ -perspective, you might reflect: *given that I believe that φ is true*, I ought to believe that ψ is true. You might also reflect: *given that I do not believe that ψ is true*, I ought not to believe that φ is true. The overall result of these reflections is a pull towards logical closure that can generate the idea that Single-Premise Closure governs thought.

My simple error theory, then, improves on its predecessor. But you might still worry about the new theory. Let me respond to a couple of anticipated concerns with it.

Objection. Suppose that you believe that φ is true and you believe that $\neg\varphi$ is true. Inside the φ -perspective, you believe that reason requires you not to believe that $\neg\varphi$ is true, but also believe that reason requires you to believe that φ is true. Similarly inside the $\neg\varphi$ -perspective. Since you would satisfy as many (or more) believed requirements by retaining the contradiction in your beliefs as you would by revising one or both of the conflicting beliefs, Satisfaction_N cannot be explained in terms of these requirements.

Reply. If you revise your belief that φ is true, then you no longer satisfy the believed requirement inside the φ -perspective to believe that φ is true, but you come to satisfy the believed requirement inside the $\neg\varphi$ -perspective not to believe that φ is true. The net gain in *satisfaction points* is zero. However, you previously seemed to violate a requirement inside the φ -perspective not to believe that $\neg\varphi$ is true and you no longer identify with this perspective. So the net gain in satisfaction points can seem to be one.

In any case, this accounting objection arises by considering believed requirements across perspectives. Inside the $\neg\varphi$ -perspective alone, the intuition that you satisfy some requirement by revising your belief that φ is true will not be offset or weakened by a contrasting intuition that in revising this belief you no longer satisfy a requirement that you had previously satisfied.

Objection. When you activate full beliefs that conflict with your other doxastic attitudes, violation and satisfaction claims can be explained within the first-person deliberative standpoint by what the evidence seemingly requires. But what if none of the objectionable beliefs are activated in the first place? Moreover, violation and satisfaction claims cannot be explained by my theory from the second-person standpoint of

advice.

Reply. These violation and satisfaction claims can still be explained by genuine evidential norms based on an advisee's or advisor's evidence and second-order requirements to self-monitor (when in force)—my error theory *supplements* rather than *replaces* Kolodny's original error theory. Of course, another possible explanation is just this: having bought into the idea that requirements like Non-Contradiction and Single-Premise Closure are normative for thought as a result of normative pressures felt in much of our own deliberations, we have extended this myth to all cases and incorporated it into our practice of giving advice.

6.10 Conclusion

In earlier chapters, I called into question the well-entrenched idea that logically valid arguments preserve truth by virtue of their logical form. There I argued that by adopting the informational view of logic, we can still hold onto something close to the standard Fregean picture. In this chapter, I challenged another popular thesis about logic—*viz.*, that rational requirements of logical coherence *as such* are normative for thought. But here my conclusion is more discouraging. In light of the Problem of Normativity in §6.2 and my error theory in §6.7 and §6.8, I submit that we should be very skeptical about the idea that logical principles like Non-Contradiction and Single-Premise Closure are genuine deliberative requirements.

I do not want to overstate things. My conclusion is not that these requirements do not exist. My target has been the idea that rational requirements like Non-Contradiction and Single-Premise Closure apply within the first-person and second-person deliberative standpoints, not the idea that these coherence requirements apply within the third-person standpoint of appraisal. If we cannot make sense of Non-Contradiction and Single-Premise Closure as deliberative norms, we might still make sense of these requirements as evaluative standards for belief-forming processes (see n. 13). Nor is my conclusion that logical validity plays no normative role in reasoning. I agree with Kolodny that, at the very least, logic governs belief *indirectly*. Relations of logical consistency and consequence inform the evidential support relation $e : S_{\mathcal{L}} \mapsto \mathbb{R}$ which in turn partly determines what reason requires, permits, and forbids you to believe. However, whereas Kolodny's error theory appeals to the patterns of evidential norms R_1 and R_2 , my theory appeals only to what agents with logically incoherent beliefs will take their evidence to require inside different categorical perspectives.

Chapter 7

Conclusion

How to teach logic? In the first lesson of an introductory logic course, I like to tell students that logically valid arguments have *good flow*. A valid argument is like a river that does not have a factory or a pig farm contaminating its water at some point between its source and estuary. When the headwaters are fresh and clean, the water will stay clean as it flows downriver. If I were to stick to the beaten path, of course, I would cash this out in terms of truth: the clean headwaters are the true premises of an argument, and good flow is the preservation of truth from the argument's premises to conclusion.

In preceding chapters, however, I have argued that this picture isn't quite right. We can hold onto the river metaphor, if we'd like. But if we want to avoid misleading our students, then we should rather cash things out in terms of incorporation: the clean headwaters are premises jointly incorporated by some body of information, and good flow is the preservation of incorporation from premises to conclusion.

Pedagogically, I suppose that it might be preferable to begin with truth preservation. After all, truth-at-a-world is conceptually simpler than incorporation-by-an-information-state (indeed, one must grasp the former notion in order to grasp the latter notion). Moreover, the truth preservation and informational views of logic deliver the same verdicts on arguments in the simple languages without informational constants that we typically work with in a first logic course. So we still get the extension of logical validity right.

However, it is customary to follow up an introductory logic course with either a course in metalogic—we have students prove soundness and completeness theorems, the Löwenheim-Skolem theorem, and so forth—or a course that surveys different logical systems—intuitionistic logic, fuzzy logic, different modal logics, and so forth—but stays within the truth preservation paradigm by defining formal notions of validity for these systems in terms of truth-in-a-model.

Let us instead set matters straight. After introducing students to logic in the usual fashion, I recommend that we quickly emphasize that the truth preservation view is only an instructive first pass. Before much time passes, we should teach that a different conception of logic as a descriptive science concerned with the structural features of information is better-suited to explain good argumentative flow in richer languages.

Within my informational framework, there is obviously much more to say about logic, good deduction, and the normative role of logic in our epistemic and linguistic practices. I close by briefly mentioning a few directions for further research.

Deductive Inquiry. Together with the informational view of logic, I have endorsed an informational conception of deductive inquiry that explains the greater variety of good argument forms in a rich language with informational constants. It is hard to see how we could make sense of good argumentation involving hypothetical reasoning, in particular, without shifting to this picture where inquirers aim to determine what is so according to information meeting the structural constraints imposed by the premises of an argument.¹

This informational view of deduction can be further developed. In Appendix A, I discuss how we can make sense of, and should make sense of, two fundamentally different kinds of supposition within this framework. Furthermore, logicians have developed natural deduction proof systems appropriate to the standard truth preservation view of logic that accurately codify good deductive argumentation in simple languages without informational constants. It would be nice to have a formal proof system that models, *inter alia*, the different forms of indirect proof and constructive dilemma discussed in Chapter 5. But I think that progress can be made. For a large fragment of the language \mathcal{L} that I have been considering, I present such a proof system, **Info**, in Appendix B.

Probability Operators. So far, I have concentrated on deductive arguments involving informational necessity and possibility operators. But there are other informational modal operators besides—for instance, *probability operators* like ‘at least as likely as’ and ‘probably’ (cf. Yalcin [2010]):

(P1) Professor Plum is at least as likely as Reverend Green to have done it.

(P2) Reverend Green is at least as likely as not to have done it.

¹In §5.1, I focused on *reductio ad absurdum* and constructive dilemma. However, as will be clear from the appendices, conditional proof should also be understood along informational lines.

- (C) Professor Plum is at least as likely as not to have done it.
- (P1) If Miss Scarlett did it, then the murder took place in the lounge.
- (P2) Miss Scarlett probably did it.
- (C) The murder probably took place in the lounge.

The informational view of logic and deductive inquiry can be extended to cover probability operators by placing quantitative measures on Boolean algebras of subsets of \mathcal{W} .² In future work, I also plan to extend the proof system **Info** to model probabilistic argumentation.

Interrogatives. The informational view can also be extended to cover *interrogative* sentences:

- (1) Did Mrs. White do it?
- (2) If Mrs. White is the murderess, did she use the rope?

As I defined in §3.3.2 the formal notion of incorporation relating an information state $i \in 2^{\mathcal{W}}$ and a declarative sentence $\varphi \in S_{\mathcal{L}}$, I define in §A.4 a formal notion of *settlement* relating information states and some interrogative sentences. We can also introduce a notion of logical consequence for interrogatives that necessarily preserves settlement by virtue of logical form (indeed, we might introduce a logical consequence relation over a language including both declarative and interrogative sentences that preserves a hybrid notion of incorporation-settlement).

Formal Theories of Incorporation. As logicians and philosophers since Tarski [1936b] and Kripke [1975] have been busy developing formal truth theories, we can get to work developing formal theories of incorporation. This is particularly pressing given Field’s argument against the truth preservation view mentioned in Chapter 4, n. 2. Adding to truth theory T , formulated in a language with an untyped truth predicate $Tr(x)$, either the sentence saying that all of T ’s axioms are true or the sentence saying that all of T ’s rules of inference preserve truth results in inconsistency. Field concludes that if we wish to maintain that logic lines up with good deductive inference such that these axioms/rules are logical truths/logically valid, then we are not in a position to consistently maintain the standard truth preservation view of logic.

Does a variant of this argument also tell against the informational view? “In light of our best incorporation theories,” a Fieldian dissenter might similarly argue, “we cannot consistently maintain that logically valid arguments preserve incorporation.” However, I think this attack

²Yalcin [2010] uses *probability measures*, but Holliday and Icard [*ms.*] establish that weaker *fuzzy measures* and a purely qualitative semantics can also handle the linguistic data.

can be resisted. Many of the truth principles that drive Field's original truth-theoretic argument, such as the intersubstitutivity of $Tr(\langle\varphi\rangle)$ and φ in transparent (non-quotational, etc.) contexts, do not have plausible information-theoretic analogues. Complications arise, however, when considering languages with both a truth predicate and informational modal operators.

Appendix A

Two Kinds of Supposition

In this first appendix, I want to pick up my discussion of supposition from §5.1. In §A.1 and §A.2, I argue that to make sense of conditional and indirect proof in rich languages with informational modal operators and indicative conditionals, we must acknowledge two distinct kinds of supposition: *lossy supposition* triggers hypothetical contexts in which the premises of an argument can fail to hold; *lossless* supposition always triggers hypothetical contexts in which one’s information incorporates everything that was incorporated before.¹ In §A.3 and §A.4, I tease out some implications of this distinction for the semantics of both indicative conditional declarative and interrogative sentences.

A.1 Lossy Supposition

Walking in the woods, you spot a creature up in a tree and reason as follows:

| | | |
|---|---|----------------------|
| 1 | The animal might be five-toed | Premise ² |
| 2 | If the animal is a sloth, then it’s two or three-toed | Premise |
| 3 | The animal has at least three toes | Premise |
| 4 | The animal is a sloth | Supposition |
| 5 | The animal is two or three-toed | From 2,4 |
| 6 | The animal is three-toed | From 3,5 |
| 7 | If the animal is a sloth, then it’s three-toed | From 4-6 |

¹I am grateful to Justin Vlastis and Seth Yalcin for suggesting this terminology.

²This first premise isn’t used in your argumentation but it will be relevant to later discussion.

This kind of hypothetical deliberation is common. To establish that certain things are so—that if the animal is a sloth then it’s three-toed—we often suppose that certain other things are so—that it’s a sloth. But how exactly does this reasoning work? In particular, what is the role of supposition?

At a high level of abstraction, the opening conditional proof works as follows. Your initial information might leave open the possibility, say, that the animal is a five-toed monkey. However, this information rules out the possibility that the animal is a sloth but is neither two-toed nor three-toed, and also the possibility that it has fewer than three toes. In supposing that the animal is a sloth, you then provisionally *add* the information that it’s a sloth to your initial information; you enter a hypothetical context in which your stronger information also rules out the possibility that the animal isn’t a sloth. Since this updated information must rule out the possibility that the sloth is a two-toed *Choloepus*, you do well to infer that if the animal is a sloth then it’s a three-toed *Bradypus*.

To make this more precise, let sentence letter S abbreviate ‘The animal is a sloth,’ natural number $n \in \mathbb{N}$ abbreviate ‘The animal is n -toed,’ and $\geq n$ abbreviate ‘The animal has at least n toes’:

$$\begin{aligned} \llbracket S \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } \text{The animal is a sloth in } w \\ \llbracket n \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } \text{The animal is } n\text{-toed in } w \\ \llbracket \geq n \rrbracket_{\mathcal{M}}^{w,i} = T & \text{ iff } \llbracket 1 \rrbracket_{\mathcal{M}}^{w,i} = \dots = \llbracket n-1 \rrbracket_{\mathcal{M}}^{w,i} = F \end{aligned}$$

Your initial information can be modeled by an information state i_1^* that incorporates the three premises: $i_1^* \triangleright \diamond 5$, $i_1^* \triangleright S \Rightarrow (2 \vee 3)$, and $i_1^* \triangleright \geq 3$. That is, i_1^* has the following three structural features depicted in Fig 5:³

- F1. Some member is a 5-world or the state is empty.
- F2. Every member that is an S -world is either a 2-world or 3-world.
- F3. No member is a 0-world, 1-world, or 2-world.

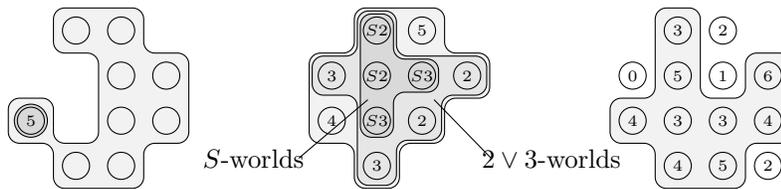


Fig 5. Information states with F1, F2, and F3

³Where $\llbracket S \rrbracket_{\mathcal{M}}$ and $\llbracket n \rrbracket_{\mathcal{M}}$ designate the set of S -worlds and n -worlds respectively in \mathcal{W} , the three features are F1. $i \cap \llbracket 5 \rrbracket_{\mathcal{M}} \neq \emptyset \vee i = \emptyset$, F2. $i \cap \llbracket S \rrbracket_{\mathcal{M}} \subseteq \llbracket 2 \rrbracket_{\mathcal{M}} \cup \llbracket 3 \rrbracket_{\mathcal{M}}$, and F3. $i \cap (\llbracket 0 \rrbracket_{\mathcal{M}} \cup \llbracket 1 \rrbracket_{\mathcal{M}} \cup \llbracket 2 \rrbracket_{\mathcal{M}}) = \emptyset$.

In the hypothetical context triggered by the supposition that the animal is a sloth, your updated information can then be explicated by the information state $i_1^* + S \subseteq i_1^*$ obtained from i_1^* by keeping only the S -worlds. By definition, $i_1^* + S \triangleright S$. In addition to having F2 and F3 which are preserved under subsets of information states—if i instantiates either property, then any subset $i' \subset i$ does as well— $i_1^* + S$ also has this structural feature depicted in Fig 6:

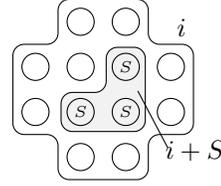


Fig 6. Information state with F4

F4. Every member is an S -world.⁴

This captures the intuitive idea that in adding the information that the animal is a sloth to your initial information, you make the minimal change required to accommodate this supposition.

Since $i_1^* + S$ has features F2 and F4, $i_1^* + S \triangleright 2 \vee 3$. Since $i_1^* + S$ has feature F3, $i_1^* + S \triangleright 3$. According to your updated information in the subproof, the animal is a three-toed *Bradypus*. But the ultimate purpose of your hypothetical reasoning is to establish what is so according to your initial information modeled by i_1^* . Back in the main categorical context of the conditional proof, your reasoning establishes that every S -world is a 3-world: $i_1^* \triangleright S \Rightarrow 3$ by the semantic clause for the indicative. According to your initial information, if the animal is a sloth then it's three-toed.

By contrast, F1 is *not* preserved under subsets of information states. The corresponding premise $\diamond 5$ is *non-persistent* (recall Def 13 in §5.1): $\exists i, i'$ where $i' \subset i$, $i \triangleright \diamond 5$, but $i' \not\triangleright \diamond 5$. Good thing, then, that $\diamond 5$ isn't used in the subproof triggered by your supposition which is of the following kind:

Def 15. A supposition is lossy just in case for some information state i and sentence φ where $i \triangleright \varphi$, if one's initial information can be modeled by state i and one makes this supposition, then one thereby enters a hypothetical context in which one's information can be modeled by a state i' where $i' \not\triangleright \varphi$.

If i_1^* contains an S -world, then $i_1^* + S$ is nonempty and excludes 5-worlds. Thus, importing $\diamond 5$ into the subproof would lead to trouble:

⁴This feature is F4. $i \subseteq [S]_{\mathcal{M}}$.

| | | |
|----|---|-------------|
| 1 | The animal might be five-toed | Premise |
| 2 | If the animal is a sloth, then it's two or three-toed | Premise |
| 3 | The animal has at least three toes | Premise |
| 4 | The animal is a sloth | Supposition |
| 5 | The animal is two or three-toed | From 2,4 |
| 6 | The animal is three-toed | From 3,5 |
| 7 | It's not the case that the animal might be five-toed | From 6 |
| 8 | \perp | From 1,7 |
| 9 | The animal is fifty-toed | From 8 |
| 10 | If the animal is a sloth, then it's fifty-toed | From 4-9 |

Bringing this non-persistent premise into the subproof at step 8, you infer that if the animal is a sloth then it's fifty-toed—a bizarre conclusion if this conditional has its usual meaning.⁵

A.2 Lossless Supposition

Given the presence of informational constants—and so non-persistent sentences—in our language, it might seem that all supposition is lossy and we must always exercise caution when importing certain kinds of sentences into subproofs.⁶ In fact, there are also *lossless* suppositions.

⁵The reason for this qualification will emerge in §A.3.

⁶So far, I've been discussing what Joyce [1999] calls “indicative”/“matter-of-fact” supposition. However, if your initial information had ruled out the possibility that the animal is a sloth, then you might still have investigated what *would* be so if the animal *were* a sloth. In the hypothetical context induced by a “subjunctive” supposition that the animal is a sloth, a reasoner investigates what is so according to information incorporating that the animal is a sloth but failing to subsume all of her initial information, some of which must be temporarily abandoned to preserve consistency. Determining the post-supposition information state i' from the pre-supposition information state i requires additional structure. For instance, we might employ a similarity ranking of the worlds in \mathcal{W} (cf. Stalnaker [1968], Lewis [1973]) and consider close worlds in which the animal is a sloth. Irrespective of the exact details, though, we can say at least this much: $i \cap i' = \emptyset$ and $i' \triangleright S$. Presumably also $i' \triangleright S \Rightarrow (2 \vee 3)$, so $i' \not\triangleright \diamond 5$.

Note that subjunctive supposition is even more lossy than indicative supposition. After an indicative supposition, only non-persistent sentences can fail. However, after a subjunctive supposition, persistent sentences like ≥ 3 might no longer hold.

To see this, consider the following *reductio*:

| | | |
|---|---|-------------|
| 1 | The animal might be five-toed | Premise |
| 2 | If the animal is a sloth, then it's two or three-toed | Premise |
| 3 | <div style="border-left: 1px solid black; padding-left: 10px;">The animal must be a sloth</div> | Supposition |
| 4 | <div style="border-left: 1px solid black; padding-left: 10px;">The animal is a sloth</div> | From 3 |
| 5 | The animal is two or three-toed | From 2,4 |
| 6 | The animal might be neither two nor three-toed | From 1 |
| 7 | \perp | From 5,6 |
| 8 | The animal might not be a sloth | From 3-7 |

This argumentation is impeccable.⁷ How exactly does it work? As before, your initial information i_2^* incorporates the premises: $i_2^* \triangleright \diamond 5$ and $i_2^* \triangleright S \Rightarrow (2 \vee 3)$. Assume also that $i_2^* + S \neq \emptyset$. It is tempting to then say that your information in the hypothetical context can be modeled by this nonempty information state $i_2^* + S$. After your supposition that the animal must be a sloth, you attend to the closest hypothetical extension of your initial information that accommodates this supposition.

This would, however, be a mistake. $i_2^* + S$ has features F2 and F4, so $i_2^* + S \triangleright 2 \vee 3$. But then $i_2^* + S$ *cannot* have feature F1—none of its members are 5-worlds. If $i_2^* + S$ explicated your updated information in the hypothetical context, then it would be erroneous to import $\diamond 5$ into the subproof. It would be erroneous to reflect inside the hypothetical context that the animal might be neither two-toed nor three-toed. However, the *reductio* is fine, and therefore your updated information should *not* be explicated by $i_2^* + S$.

So what are you doing when you suppose that the animal must be a sloth? How to model your information in the hypothetical context if not by $i_2^* + S$? There is an alternative account that makes good sense of the *reductio*. In supposing that the animal must be a sloth, you temporarily investigate a body of information that rules out the possibility that it isn't a sloth but also still has the structural features corresponding to each of the argument's premises. Formally, your information should be explicated by an information state $i_2^* \oplus S$ that is just i_2^* if $i_2^* \triangleright S$ and \emptyset otherwise (more generally, $i \oplus \varphi = i$ if $i \triangleright \varphi$ and $i \oplus \varphi = \emptyset$ otherwise).

⁷Indeed, as discussed in §5.1, I think that your argumentation is also impeccable if you suppose instead that the animal *is* a sloth and reason to the conclusion that it might not be a sloth.

This information state clearly has features F1, F2, *and* F4. By the end of the subproof, of course, you come to recognize that $i_2^* \oplus S = \emptyset$ —no nonempty information state has all three features. Back in the main categorical context of the proof where i_2^* has both F1 and F2, then, you establish that i_2^* cannot have F4— $i_2^* \not\triangleright S$, so $i_2^* \triangleright \diamond\neg S$. According to your initial information, the animal might not be a sloth.

Your supposition is of the following kind:

Def 16. A supposition is lossless just in case for every information state i and sentence φ , if one’s initial information can be modeled by state i where $i \triangleright \varphi$ and one makes this supposition, then one thereby enters a hypothetical context in which one’s information can be modeled by state i' where $i' \triangleright \varphi$.

In the conditional proof in §A.1, the non-persistent premise $\diamond 5$ fails to hold after your lossy supposition. In the indirect proof, by contrast, there is no loss.

Notice how my informational account of the *reductio* parallels the standard truth-centric account. On the old story found in virtually any logic textbook, a reasoner begins her inquiry with premises $\varphi_1, \dots, \varphi_n$ that constrain the way the world might be. If φ_i is true, then the world satisfies the constraint corresponding to this premise. After supposing that ψ is true, the reasoner then establishes that premises $\varphi_1, \dots, \varphi_n$ plus ψ cannot jointly be true—no world meets all of the corresponding constraints—and concludes from the truth of the premises that $\neg\psi$ is true.

On my informational account, by comparison, you begin inquiry with premises $\diamond 5$ and $S \Rightarrow (2 \vee 3)$, each of which corresponds to a constraint on information states—*viz.*, $\triangleright \diamond 5$ and $\triangleright S \Rightarrow (2 \vee 3)$. After supposing that the animal must be a sloth, you then establish that no nonempty information state satisfies these structural constraints and also satisfies the constraint $\triangleright S$ imposed by your supposition. Since your initial body of information has the structure corresponding to the premises, you conclude that this information cannot have the structure corresponding to the supposition.

A.3 Conditional Declaratives

Now, I have argued that we need both lossy and lossless supposition to make sense of both conditional and indirect proof in languages that include informational constants. For conditional proof, we need lossy

supposition. For indirect proof, we need lossless supposition.⁸ Let me now turn to some implications of this distinction.

The first implication concerns the semantics of indicative conditional declarative sentences like these:

- (1) If the animal is a sloth, then it's five-toed.
- (2) If the animal must be a sloth, then it's five-toed.

On the Yalcin-Kolodny-MacFarlane (YKM) semantics in Chapter 2, these sentences have the same semantic value. Our discussion in §A.1 and §A.2 thus suggests that the YKM semantics is either incorrect or incomplete. Let me explain.

Philosophers, linguists, and psychologists alike have stressed that there is an intimate connection between suppositions and indicative conditionals. The *locus classicus* is this famous footnote in Ramsey's *General Propositions and Causality* [1929]:⁹

If two people are arguing 'If p will q ?' and are both in doubt as to p , they are adding p hypothetically to their stock of knowledge and arguing on that basis about q ; so that in a sense 'If p , q ' and 'If p , \bar{q} ' are contradictories. (p. 143)

This suggests a two-stage test for determining whether an indicative conditional holds: first, suppose that its antecedent holds, and second, check if its consequent holds in the triggered hypothetical context. If the consequent does hold, then the entire conditional holds in the categorical context.

I hope this rings a bell. There are many different versions of the *Ramsey Test* floating around the literature.¹⁰ But on one interpretation I favor, the Ramsey Test is just *conditional proof* understood along informational lines. Recall that in the conditional proof in §A.1, you begin by tentatively adding the information that the animal is a sloth to

⁸This is to say neither that *all* conditional proofs involve lossy supposition, nor that *all* indirect proofs involve lossless supposition. It is to say only that some conditional proofs certainly involve lossy supposition and some indirect proofs involve lossless supposition. More on this in a moment.

⁹See also Stalnaker [1968], Mackie [1973], Adams [1975], Edgington [1986], Bennett [2003], Evans and Over [2004], Bradley [2007], Barnett [2008], Isaacs and Rawlins [2008], Leitgeb [2011], and many others.

¹⁰Ramsey's footnote continues: "We can say that they are fixing their degrees of belief in q given p ." So there are also quantitative versions of the Ramsey Test—for instance, Adams' [1975] famous thesis that the 'probability' of an indicative equals the conditional probability of its consequent given its antecedent, provided that the probability of its antecedent is nonzero.

your initial information.¹¹ After this indicative supposition, you come to recognize that your updated information incorporates that the animal is three-toed, so you do well to infer that if the animal is a sloth then it's a *Bradypus*.¹²

Moreover, the Ramsey Test—that is, conditional proof—is *built into* the YKM semantics. The method for establishing that indicatives hold and their semantic values go hand in hand. Recall this semantic clause:

$$\llbracket \varphi \Rightarrow \psi \rrbracket_{\mathcal{M}}^{w,i} = T \quad \text{iff} \quad i + \varphi \triangleright \psi$$

The evaluation procedure for $\varphi \Rightarrow \psi$ can also be regarded as a two-stage affair that mirrors conditional proof with lossy supposition: first, we find the maximal subset $i + \varphi \subseteq i$ that incorporates the antecedent φ , and second, we check whether this updated state incorporates the consequent ψ . The indicative conditional is true at $\langle w, i \rangle$ just in case this test passes.

For example, the clauses for our sample sentences are these:

$$\begin{aligned} \llbracket S \Rightarrow 5 \rrbracket_{\mathcal{M}}^{w,i} = T & \quad \text{iff} \quad i + S \triangleright 5 \\ \llbracket \Box S \Rightarrow 5 \rrbracket_{\mathcal{M}}^{w,i} = T & \quad \text{iff} \quad i + \Box S \triangleright 5 \end{aligned}$$

To evaluate (1) at $\langle w, i \rangle$, we first update i with the information in the antecedent to obtain $i + S$ and then check whether this updated state consists solely of 5-worlds. To evaluate (2), we can follow the exact same procedure—since $i \triangleright S$ if and only if $i \triangleright \Box S$, $i + S = i + \Box S$, so the semantic values of (1) and (2) are identical.

The YKM semantics has much to recommend it. For one thing, it provides an elegant explanation of why indicative conditionals whose antecedents are “epistemic contradictions” sound defective (cf. Yalcin [2007]):

¹¹Unlike Ramsey, though, I don't want to insist that your initial information is what you know. Two people might argue ‘If p will q ?’ according to information that is neither known by either party nor distributed knowledge.

¹²As Bradley [2007] notes, the Ramsey Test is best regarded as a schema for determining whether a variety of natural language conditionals hold:

It is commonly observed that there is more than one kind of ordinary language conditional, although exactly how to classify the various kinds is a matter of some dispute. One advantage of the Ramsey Test hypothesis is that it allows us to link this observation to the fact that there are different kinds of suppositions or ways of supposing something true. For the Ramsey Test can be treated as a test *schema* with different types of belief revision being suitable for testing the credibility of different kinds of conditionals. (p. X)

Input indicative supposition into the test schema and we get conditional proof for indicative conditionals. Input subjunctive supposition and we get conditional proof for counterfactuals.

- (3) # If the animal is a sloth and it might not be a sloth, then it's five-toed.
- (4) # If the animal might not be a sloth and it's a sloth, then it's five-toed.

$i + S \wedge \diamond \neg S = i + \diamond \neg S \wedge S = \emptyset$, so the evaluation process for these conditionals has us check whether their consequents are incorporated by the empty set \emptyset —a rather odd test.

Nevertheless, there is reason to worry. The equivalence of (1) and (2) is troubling in light of our previous discussion of supposition. Again: on the YKM semantics, the semantic values of *all* indicative conditionals embed the Ramsey Test with *lossy* supposition. But there is also lossless supposition. In §A.1, your supposition that the animal is a sloth is lossy. In §A.2, your supposition that the animal must be a sloth is lossless. Preserving the bond between suppositions and indicative conditionals seemingly requires that the semantic values of (1) and (2) come apart.

Consider the following conditional proof:

| | | |
|---|---|-------------|
| 1 | The animal might be a monkey | Premise |
| 2 | The animal must be a sloth | Supposition |
| 3 | It's not the case that the animal might be a monkey | From 2 |
| 4 | \perp | From 1,3 |
| 5 | The animal is fifty-toed | From 4 |
| 6 | If the animal must be a sloth, then it's fifty-toed | From 2-5 |

If you add the information that the animal is a sloth to your initial information at step 2, then you shouldn't import the premise into the subproof at step 4. However, if your supposition here is lossless, then importing the premise is fine, and the final step 6 seems to be a felicitous use of conditional proof based on your lossless supposition. But then the conclusion cannot have the YKM semantics on which this conditional does not follow from the premise that the animal might be a monkey.¹³

The Ramsey Test with lossless supposition suggests this alternative clause for (2):

$$\llbracket \Box S \Rightarrow 5 \rrbracket_{\mathcal{M}}^{w,i} = T \quad \text{iff} \quad i \oplus S \triangleright 5^{14}$$

¹³The conclusion doesn't follow on *any* of the consequence relations introduced in Chapters 2 and 3.

¹⁴Notice that on this semantics, $\llbracket \Box S \Rightarrow 5 \rrbracket_{\mathcal{M}}^{w,i} = \llbracket \Box S \supset \Box 5 \rrbracket_{\mathcal{M}}^{w,i}$.

Semantically evaluating (2) at $\langle w, i \rangle$ now involves the following two-stage procedure. First, we check whether $i \triangleright S$. If this initial test fails and so $i \oplus S = \emptyset$ —as it does when i includes an M -world in which the animal is a monkey—then (2) is true at $\langle w, i \rangle$. After all, if i explicates your initial body of information, then \emptyset explicates your information after a lossless supposition that the animal must be a sloth, and the empty set incorporates everything. If the first test passes and so $i \oplus S = i$, we then check whether $i \triangleright 5$. At this point, (2) is true at $\langle w, i \rangle$ if and only if this second test passes. After all, i still explicates your information after the lossless supposition, so (2) holds according to your initial information if and only if its consequent holds according to this same body of information.

It is tempting to conclude that the YKM semantics is incorrect. Indeed, one might plump for this general clause that agrees with the YKM semantics for (1) but agrees with the semantics for (2) based on lossless supposition:

$$\llbracket \varphi \Rightarrow \psi \rrbracket_{\mathcal{M}}^{w,i} = T \quad \text{iff} \quad i \boxplus \varphi \triangleright \psi$$

where, recall from §3.2, that $i \boxplus \varphi = i \cap \{w : \llbracket \varphi \rrbracket_{\mathcal{M}}^{w,i} = T\}$.¹⁵ By doing this one cannot explain the defectiveness of (3) and (4) as before. But perhaps the embedded epistemic contradiction data can be satisfactorily explained in some other way.¹⁶

However, there is an alternative conclusion—namely, that the YKM semantics is not incorrect but *incomplete*. Throughout this appendix, I have assumed that the non-modal supposition that the animal is a sloth is lossy whereas the modal supposition that the animal must be a sloth is lossless. But plausibly you can make and attribute both lossy and lossless suppositions with either modal or non-modal forms. Sometimes supposing that the animal is a sloth is lossless like the modal supposition in §A.2. Sometimes supposing that the animal must be a sloth is lossy like the non-modal supposition in §A.1.

If this is right, then the link between suppositions and indicatives suggests that (1) and (2) are ambiguous. The YKM semantics covers their usual interpretation based on lossy supposition. But indicatives

¹⁵It is easy to verify that $i \boxplus S = i + S$ and $i \boxplus \Box S = i \oplus S$. I suspect that this semantic clause will appeal to fans of Veltman’s [1996] dynamic *update semantics* on which the meaning of an expression is its *context change potential* (CCP)—a function $\llbracket \]$ from information states to information states. In particular, $i \llbracket S \rrbracket i + S$ and $i \llbracket \Box S \rrbracket i \oplus S$.

¹⁶Interestingly, Yalcin also endorses a semantics along these lines in the appendix of [2012a] but he isn’t concerned in this paper with indicative conditionals that embed epistemic contradictions.

also have this alternative semantic clause based on lossless supposition:

$$\llbracket \varphi \Rightarrow \psi \rrbracket_{\mathcal{M}}^{w,i} = T \quad \text{iff} \quad i \oplus \varphi \triangleright \psi^{17}$$

Admittedly, a lossless kind of indicative conditional is something of a curiosity. It is strictly weaker than the standard lossy indicative.¹⁸ In fact, a lossless indicative is so weak that it is hard to find environments in which we would want to use it. I suppose that you might talk or think through a *reductio* as follows:

Suppose that the animal is/must be a sloth. Then it's either two-toed or three-toed. But it might be five-toed. Contradiction. If the animal is/must be a sloth, then $0=1$. Thus, the animal might not be a sloth.¹⁹

But even here the conditional 'If the animal is/must be a sloth then $0=1$ ' is redundant. You can conclude that the animal might not be a sloth without first inferring this conditional.

It is ultimately an empirical question whether the natural language indicative conditional has the single semantic clause involving \boxplus or the two clauses involving $+$ and \oplus . I will not attempt to settle it here. I am content to conclude that the Ramseyan correspondence between suppositions and indicatives pushes us towards these different options.

A.4 Conditional Interrogatives

The second implication concerns indicative conditional interrogatives that amalgamate if-clauses and polar interrogatives in the indicative mood:

(5) If the animal is a sloth, is it five-toed?

(6) If the animal must be a sloth, is it five-toed?

As you might already suspect, the implication is this: the distinction between lossy and lossless supposition suggests that (5) and (6) either have different semantics or these interrogatives are both ambiguous. In the rest of this paper, let us flesh out the second option and consider how one's response to (5) might depend on how its ambiguity is resolved.

On its more natural reading, this interrogative raises the issue of whether the animal is five-toed on the *lossy* supposition that it's a sloth. To address (5), one can follow Ramsey's suggestion and first tentatively

¹⁷On this semantics, $\llbracket \varphi \Rightarrow \psi \rrbracket_{\mathcal{M}}^{w,i} = \llbracket \Box \varphi \supset \Box \psi \rrbracket_{\mathcal{M}}^{w,i}$.

¹⁸If \Rightarrow_1 and \Rightarrow_2 designate the lossy and lossless indicative conditional respectively, then $\varphi \Rightarrow_1 \psi \models_I \varphi \Rightarrow_2 \psi$ but $\varphi \Rightarrow_2 \psi \not\models_I \varphi \Rightarrow_1 \psi$.

¹⁹Notice that $S \Rightarrow_2 \perp \models_I \Diamond \neg S$ and $S \Rightarrow_1 \perp \models_I \neg S$.

add the information that the animal is a sloth to one's initial body of information.²⁰ In the induced hypothetical context, one should then consider the question expressed by 'Is the animal five-toed?' and answer it on the basis of this updated information. One's response to the original conditional interrogative should be the same as one's response to this unconditional interrogative.

This procedure, of course, is just conditional proof for the kind of lossy indicative conditional declaratives that are candidate responses to (5). If your initial information can be explicated by i_2^* from §A.2, then your response should be something like this:

- (7) No. If the animal is a sloth, then it isn't five-toed. It's either two-toed or three-toed.

The issue of whether the animal is five-toed is unresolved relative to your initial body of information since this leaves open the possibilities both that the animal is five-toed and that the animal isn't five-toed. But this issue is resolved relative to your updated information inside the hypothetical context, explicated by $i_2^* + S$, since this information rules out the possibility that the animal is a sloth but is neither two-toed nor three-toed. The remaining possibilities are all ones in which the animal isn't five-toed.

On its less natural alternative reading, (5) raises the issue of whether the animal is five-toed on the *lossless* supposition that the animal is a sloth. This is the kind of interrogative that one might utter before embarking on a *reductio* proof. To address (5) on this reading, one should first check whether one's initial information rules out the possibility that the animal isn't a sloth. If this test fails, respond 'yes *and* no' since anything goes according to one's degenerate information after the lossless supposition. If this test passes and so one's salient information remains unchanged after the supposition, one should next consider the question expressed by 'Is the animal five-toed?' and respond to this unconditional interrogative and the original conditional interrogative in the same way.

In particular, if i_2^* models your initial information, then your response should be something like this:

- (8) Yes and no. If the animal is a sloth, then anything goes. The animal is five-toed. The animal is fifty-toed for that matter.²¹

²⁰Notice that Ramsey's famous footnote in §A.3 is explicitly about conditional *interrogatives*.

²¹Admittedly, this response sounds unnatural. I leave it as an open empirical question whether conditional interrogatives like (5) admit both lossy and lossless readings.

Since $i_2^* \oplus S = \emptyset$, you should not respond with (7).

We can sharpen the alternative readings of (5) with some more formal semantics. On the *partition semantics* for questions inspired by Hamblin [1958] and later developed and employed by Groenendijk and Stokhof [1984], Hulstijn [1997], Groenendijk [1999], Velissaratou [2000], Isaacs and Rawlins [2008], Yalcin [2011], Ciardelli and Roelofsen [2011], and many others, an unconditional interrogative sentence $\varphi?$ determines a partition of \mathcal{W} , or a subregion of \mathcal{W} , into mutually exclusive cells. Each cell is a complete answer to the question expressed by $\varphi?$ and the true answer is the cell containing the actual world @. For example, the polar interrogative sentence $5?$ bipartitions the subregion of \mathcal{W} in which its presuppositions are met into one cell containing the 5-worlds and another cell containing the remaining worlds.

Let us now define a formal notion of *settlement* relating information states and unconditional interrogative sentences akin to the notion of incorporation from §3.3.2 relating information states and declarative sentences:

Def 17. $i \triangleright \varphi?$ (read: i settles $\varphi?$, or $\varphi?$ is settled by i) if and only if i is a subset of some cell of the partition determined by $\varphi?$.²²

In the special case where $\varphi?$ is polar, $i \triangleright \varphi?$ just in case either $i \triangleright \varphi$ or $i \triangleright \neg\varphi$. The interrogative sentence $\varphi?$ is settled *correctly* by i if the cell inhabited by this information state is the true answer to the question expressed by this sentence. This can occur even when $@ \notin i$ (see Fig 7).

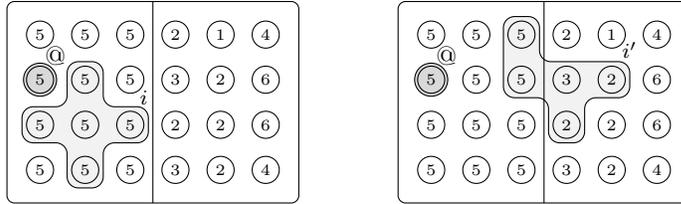


Fig 7. i settles $5?$ correctly whereas i' does not even settle $5?$

The semantics of conditional interrogatives is more controversial. Some linguists claim that $S \Rightarrow 5?$ expresses a question with a more complex answer structure than that expressed by $5?$.²³ However, if, as suggested above, to utter $S \Rightarrow 5?$ is just to raise the issue of whether the

²²I'll extend this definition to conditional interrogatives in a moment.

²³Hulstijn [1997], for instance, suggests that $S \Rightarrow 5?$ tripartitions a region of \mathcal{W} into a cell of $S \wedge 5$ -worlds, a cell of $S \wedge \neg 5$ -worlds, and a cell of $\neg S$ -worlds. Velissaratou [2000] and Ciardelli and Roelofsen [2011] abandon partitions altogether and propose that the complete answers to the question expressed by $S \Rightarrow 5?$ needn't be mutually

animal is five-toed *on the supposition that it's a sloth*, then we needn't assign a new formal semantic object to $S \Rightarrow 5?$. We can get by with the bipartition determined by $5?$ alone, so long as our semantic-pragmatic theory reflects that the bipartitioning takes effect only inside the hypothetical context triggered by the supposition that the animal is a sloth.²⁴

Now the ambiguity comes in. On the more natural reading of our sample sentence (5) based on lossy supposition, $i \triangleright S \Rightarrow 5?$ just in case $i + S \triangleright 5?$. On the less natural reading based on lossless supposition, $i \triangleright S \Rightarrow 5?$ just in case $i \oplus S \triangleright 5?$. Information state i_2^* , in particular, settles both the lossy and lossless versions of (5) since $i_2^* + S \triangleright 5?$ and $i_2^* \oplus S \triangleright 5?$. However, i_2^* settles lossy and lossless $S \Rightarrow 5?$ in different ways since $i_2^* + S$ and $i_2^* \oplus S$ settle $5?$ in different ways. As illustrated in Fig 8, $i_2^* + S \triangleright \neg 5$, so a satisfactory response to the lossy version of (5) is (7). But since $i_2^* \oplus S = \emptyset$, $i_2^* \oplus S \triangleright 5$ and $i_2^* \oplus S \triangleright \neg 5$, so a satisfactory response to the lossless version of (5) is (8).

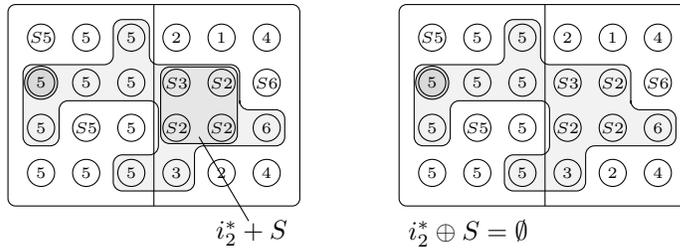


Fig 8. $i_2^* + S$ is a subset of only the right cell whereas $i_2^* \oplus S$ is a subset of both

exclusive: one complete answer is the set of $S \triangleright 5$ -worlds while the other is the set of $S \triangleright \neg 5$ -worlds. However, see Velissaritou [2000] for forceful objections to the tripartition analysis of conditional interrogatives, and Isaacs and Rawlins [2008] for objections to the non-partition analysis.

²⁴In Isaacs and Rawlin's [2008] dynamic semantics framework, for instance, the CCP of a conditional interrogative sentence can be broken into two pieces: first, the "macro-context" of a discourse is updated by the if-clause which adds a new hypothetical context to the top of the stack, and second, the embedded interrogative partitions this hypothetical context. See their paper for details.

Appendix B

The Natural Deduction System `Info`

In this second appendix, I present a Fitch-style natural deduction proof system, `Info`, appropriate to the informational view of logic. In §B.1, I define the languages relevant to this system. In §B.2, I catalog its rules. In §B.3, I prove soundness with respect to informational consequence. In §B.4, I prove completeness.

B.1 Languages

There will be a few languages in play. The most rudimentary is the language of sentential logic \mathcal{L}_0 :

Syntax of \mathcal{L}_0 . The symbols of \mathcal{L}_0 are $A, B, C, \dots, \perp, \neg, \vee, \wedge$, and parentheses.¹ The atomic sentences in $S_{\mathcal{L}_0}$ are A, B, C, \dots , and \perp . If $P, Q \in S_{\mathcal{L}_0}$, then $\neg P, (P \vee Q), (P \wedge Q) \in S_{\mathcal{L}_0}$. Nothing else is in $S_{\mathcal{L}_0}$.

Semantics of \mathcal{L}_0 . See Chapter 2.

The next language \mathcal{L}_1 is a restricted fragment of the language from Chapter 2 sans complex embeddings of \Box, \Diamond , and \Rightarrow , and certain hybrid sentences that combine subsentences involving informational constants with those not involving these constants:²

Syntax of \mathcal{L}_1 . The symbols of \mathcal{L}_1 are those of \mathcal{L}_0 plus \Box, \Diamond , and \Rightarrow .³ If $P, Q \in S_{\mathcal{L}_0}$, then $P, \Box P, \Diamond P, (P \Rightarrow Q) \in S_{\mathcal{L}_1}$. If $\varphi, \psi \in S_{\mathcal{L}_1} \setminus S_{\mathcal{L}_0}$, then $\neg\varphi, (\varphi \vee \psi), (\varphi \wedge \psi) \in S_{\mathcal{L}_1}$. Nothing else is in $S_{\mathcal{L}_1}$.

¹ \supset and \equiv can be defined in the usual fashion.

²For example, $\Diamond\Diamond A, A \Rightarrow \Box B, A \Rightarrow (B \Rightarrow C)$, and $\Diamond A \vee B$ are not in $S_{\mathcal{L}_1}$. Working with the simpler language \mathcal{L}_1 allows for a simpler proof system, and makes it easier to prove completeness in §B.4. However, simplicity comes at the cost of limited applicability. I plan to extend `Info` in future research.

³Ignoring the subtleties in §A.3, I assume here that \Rightarrow is the standard lossy indicative.

Semantics of \mathcal{L}_1 . See Chapter 2.

The final language \mathcal{L}_2 is the *language of incorporation*:

Syntax of \mathcal{L}_2 . The symbols of \mathcal{L}_2 are those of \mathcal{L}_0 plus i , $+$, and the relation \triangleright . The atomic sentences in $S_{\mathcal{L}_2}$ are all of the form $i + P \triangleright Q$ where $P, Q \in S_{\mathcal{L}_0}$ (I omit the $+P$ when P is logically equivalent to $\neg\perp$). If $\varphi, \psi \in S_{\mathcal{L}_2}$, then $\neg\varphi, (\varphi \vee \psi), (\varphi \wedge \psi) \in S_{\mathcal{L}_2}$. Nothing else is in $S_{\mathcal{L}_2}$.

Semantics of \mathcal{L}_2 . $\llbracket i + P \triangleright Q \rrbracket^{w,i} = T$ iff $\forall w' \in i + P (\llbracket Q \rrbracket^{w',i+P} = T)$ iff $i + P \triangleright Q$. For the rest, see Chapter 2.

Note that $\llbracket i + P \triangleright Q \rrbracket^{w,i} = \llbracket P \Rightarrow Q \rrbracket^{w,i}$, so \mathcal{L}_2 is not really needed in addition to \mathcal{L}_1 . But the language of incorporation will be put to good use. With it, the informational background of argumentation in \mathcal{L}_1 will be made explicit.

Consider $| \cdot | : S_{\mathcal{L}_1} \mapsto S_{\mathcal{L}_2}$ where $P, Q \in S_{\mathcal{L}_0}$ and $\varphi, \psi \in S_{\mathcal{L}_1} \setminus S_{\mathcal{L}_0}$:

$$\begin{aligned} |P| &= i \triangleright P \\ |\Box P| &= i \triangleright P \\ |\Diamond P| &= \neg i \triangleright \neg P \\ |P \Rightarrow Q| &= i + P \triangleright Q \\ |\neg\varphi| &= \neg|\varphi| \\ |\varphi \vee \psi| &= |\varphi| \vee |\psi| \\ |\varphi \wedge \psi| &= |\varphi| \wedge |\psi| \end{aligned}$$

A proof by induction on the complexity of sentences in $S_{\mathcal{L}_1}$ establishes that $i \triangleright \varphi$ if and only if $\forall w \in i (\llbracket |\varphi| \rrbracket^{w,i} = T)$.⁴ The argumentation $|\Pi|$ in \mathcal{L}_2 obtained by substituting $|\varphi|$ for φ at each line of argumentation Π in \mathcal{L}_1 then keeps track of incorporation facts as Π proceeds. For example:⁵

⁴The proof is left to the punctilious reader. Note that when $i \neq \emptyset$, $i \triangleright \varphi$ iff $\forall w (\llbracket |\varphi| \rrbracket^{w,i} = T)$. But this stronger equivalence fails in the degenerate case where $i = \emptyset$: $\emptyset \triangleright \Diamond A$, but $\llbracket |\Diamond A| \rrbracket^{w,\emptyset} = \llbracket \neg i \triangleright \neg A \rrbracket^{w,\emptyset} = F$ for any w , since $\emptyset \triangleright \neg A$.

⁵Given my discussion in Appendix A, we should really distinguish between lossy and lossless supposition in our Fitch-style systems. But for ease of exposition, I stick to the standard proof structure here.

| | | | | |
|---|----------------------------------|-------------|---|--|
| 1 | $A \Rightarrow B$ | Premise | 1 | $i + A \triangleright B$ |
| 2 | $\Box \neg B \wedge \Box \neg C$ | Premise | 2 | $i \triangleright \neg B \wedge i \triangleright \neg C$ |
| 3 | \overline{A} | Supposition | 3 | $\overline{i \triangleright A}$ |
| 4 | B | From 1,3 | 4 | $\overline{i \triangleright B}$ |
| 5 | $\Box \neg B$ | From 2 | 5 | $i \triangleright \neg B$ |
| 6 | \perp | From 4,5 | 6 | $i \triangleright \perp$ |
| 7 | $\neg A$ | From 3-6 | 7 | $i \triangleright \neg A$ |
| 8 | $\neg \Diamond A$ | From 7 | 8 | $\neg \neg i \triangleright \neg A$ |

The goodness of Π will turn on whether $|\Pi|$ is or can be expanded into a proof in the system **Info**. If $|\Pi|$ or its expansion is a proof in **Info**, then we might confer proofhood on Π as an inherited honorific status.

B.2 Rules of Info

These proof rules fix the *inner logic* of \triangleright , where $P, Q, R, S \in S_{\mathcal{L}_0}$ and $|\varphi| \vdash_s^* |\psi|$ designates that $|\psi|$ has been proven in subproof s initiated with $|\varphi|$ using only certain kinds of sentences:

(\star) Convert $|\chi| \in S_{\mathcal{L}_2}$ into *disjunctive normal form*: a sentence of the form $\chi_1 \vee \dots \vee \chi_n$ where each χ_i is a conjunction of literals—atomic sentences $i + P \triangleright Q$ or their negations $\neg i + P \triangleright Q$. If any literal is of the form $\neg i + P \triangleright Q$ and $|\chi|$ is appealed to inside of s , then $|\varphi| \not\vdash_s^* |\psi|$.

$$\begin{array}{l}
\neg\text{Intro}_{\triangleright} \quad \frac{i + P \triangleright Q \vdash_s^* i + P \triangleright \perp}{i + P \triangleright \neg Q} \\
\neg\text{Elim}_{\triangleright} \quad \frac{i + P \triangleright \neg \neg Q \quad \neg i + P \triangleright \neg \neg Q}{i + P \triangleright Q \quad \neg i + P \triangleright Q} \\
\vee\text{Intro}_{\triangleright} \quad \frac{i + P \triangleright Q}{i + P \triangleright Q \vee R} \\
\vee\text{Elim}_{\triangleright} \quad \frac{i + P \triangleright Q \vee R \quad i + P \triangleright Q \vdash_{s_1}^* i + P \triangleright S \quad i + P \triangleright R \vdash_{s_2}^* i + P \triangleright S}{i + P \triangleright S} \\
\wedge\text{Intro}_{\triangleright} \quad \frac{i + P \triangleright Q \quad i + P \triangleright R}{i + P \triangleright Q \wedge R} \\
\wedge\text{Elim}_{\triangleright} \quad \frac{i + P \triangleright Q \wedge R \quad i + P \triangleright Q \wedge R}{i + P \triangleright Q \quad i + P \triangleright R} \\
\perp\text{Intro}_{\triangleright} \quad \frac{i + P \triangleright Q \quad i + P \triangleright \neg Q}{i + P \triangleright \perp}
\end{array}$$

$$\begin{array}{l}
\perp\mathbf{Elim}_{\triangleright} \quad \frac{i + P \triangleright \perp}{i + P \triangleright Q} \\
\Diamond\mathbf{Intro}^6 \quad \frac{i + P \triangleright Q}{\neg i + P \triangleright \neg Q} \\
\Diamond\mathbf{Elim} \quad \frac{\neg i \triangleright \neg P \quad i + P \triangleright Q}{\neg i \triangleright \neg Q} \\
\Rightarrow\mathbf{Intro} \quad \frac{i \triangleright P \vdash_s^* i \triangleright Q}{i + P \triangleright Q} \\
\Rightarrow\mathbf{Elim} \quad \frac{i \triangleright P \quad i + P \triangleright Q}{i \triangleright Q}
\end{array}$$

These proof rules fix the *outer logic* of \triangleright , where $|\varphi|, |\psi|, |\chi| \in S_{\mathcal{L}_2}$ and $|\varphi| \vdash_s |\psi|$ designates that $|\psi|$ has been proven in subproof s initiated with $|\varphi|$ using any sentences:

$$\begin{array}{l}
\mathbf{Reit} \quad \frac{|\varphi|}{|\varphi|} \\
\neg\mathbf{Intro} \quad \frac{|\varphi| \vdash_s i \triangleright \perp}{\neg|\varphi|} \\
\neg\mathbf{Elim} \quad \frac{\neg\neg|\varphi|}{|\varphi|} \\
\forall\mathbf{Intro} \quad \frac{|\varphi|}{|\varphi| \vee |\psi|} \quad \frac{|\varphi|}{|\psi| \vee |\varphi|} \\
\forall\mathbf{Elim} \quad \frac{|\varphi| \vee |\psi| \quad |\varphi| \vdash_{s_1} |\chi| \quad |\psi| \vdash_{s_2} |\chi|}{|\chi|} \\
\wedge\mathbf{Intro} \quad \frac{|\varphi| \quad |\psi|}{|\varphi| \wedge |\psi|} \\
\wedge\mathbf{Elim} \quad \frac{|\varphi| \wedge |\psi|}{|\varphi|} \quad \frac{|\varphi| \wedge |\psi|}{|\psi|} \\
\perp\mathbf{Intro} \quad \frac{|\varphi| \quad \neg|\varphi|}{i \triangleright \perp} \\
\perp\mathbf{Elim} \quad \frac{i \triangleright \perp}{|\varphi|}
\end{array}$$

Note that if the inner logic rules $\neg\mathbf{Intro}_{\triangleright}$, $\forall\mathbf{Elim}_{\triangleright}$, and $\Rightarrow\mathbf{Intro}$ were formulated with \vdash_s instead of \vdash_s^* , then **Info** would deliver some horrible

⁶Though \Diamond and \Rightarrow are not in \mathcal{L}_2 , I label the remaining four rules as introduction and elimination rules for these symbols given the corresponding transitions in \mathcal{L}_1 that these rules effectively license. Since $|\Box P| = |P|$, $\Box\mathbf{Intro}$ and $\Box\mathbf{Elim}$ would be reiteration rules.

results. Consider something like the first example in §5.1:

| | | | | |
|---|-------------------|---|------------------------------------|--|
| 1 | $\Diamond\neg W$ | 1 | $\neg i \triangleright \neg\neg W$ | Premise |
| 2 | $R \Rightarrow W$ | 2 | $i + R \triangleright W$ | Premise |
| 3 | \overline{R} | 3 | $\overline{i \triangleright R}$ | Supposition |
| 4 | \overline{W} | 4 | $\overline{i \triangleright W}$ | \Rightarrow Elim: 2,3 |
| 5 | $\neg\Box W$ | 5 | $\neg i \triangleright W$ | \neg Elim $_{\triangleright}$: 1 |
| 6 | \perp | 6 | $i \triangleright \perp$ | \perp Intro: 4,5 |
| 7 | $\neg R$ | 7 | $i \triangleright \neg R$ | \neg Intro $_{\triangleright}$: 3-6 |

This *reductio* is infelicitous. However, its \mathcal{L}_2 -counterpart would be a proof in **Info** given the \vdash_s -form of $\neg\mathbf{Intro}_{\triangleright}$ together with some of the system's other rules. I argued in §5.1 that step 7 is problematic given that the non-persistent $\Diamond\neg W$ is imported into the subproof at step 5. Non-persistence is a semantic property but the syntactic condition (\star) also does the job.

In fact, (\star) is stronger than required. First, the appeal to some non-persistent sentences in some subproofs is harmless—for example, the appeal to $\Diamond A$ inside a subproof beginning with supposition A . Second, $|\varphi| \vdash_s^* |\psi|$ when a sentence like $|\Diamond W \vee \neg\Diamond W|$ is appealed to inside of s though $\Diamond W \vee \neg\Diamond W$ is a logical truth so is clearly persistent. However, I prove in §B.4 that this excess vigilance does not undermine **Info**'s completeness.

B.3 Soundness

I prove the following soundness theorem, where $\varphi_1, \dots, \varphi_n, \psi \in S_{\mathcal{L}_1}$ and $\{|\varphi_1|, \dots, |\varphi_n|\} \vdash_{\mathbf{Info}} |\psi|$ designates that $|\psi|$ is provable in **Info** from premises $|\varphi_1|, \dots, |\varphi_n|$:

Thm 1. If $\{|\varphi_1|, \dots, |\varphi_n|\} \vdash_{\mathbf{Info}} |\psi|$, then $\{\varphi_1, \dots, \varphi_n\} \models_I \psi$.

Proof: Assume $\{|\varphi_1|, \dots, |\varphi_n|\} \vdash_{\mathbf{Info}} |\psi|$ and $i \triangleright \varphi_1, \dots, i \triangleright \varphi_n$. To show that $i \triangleright \psi$, I first show that **Info**'s simpler inferential rules licensing transitions from input sentences $|\varphi_1^I|, \dots, |\varphi_n^I|$ to output sentence $|\varphi^O|$ preserve incorporation in \mathcal{L}_1 . That is, $i \triangleright \varphi^O$ if $i \triangleright \varphi_1^I \wedge \dots \wedge \varphi_n^I$. Keep in mind that $i \triangleright \varphi$ if and only if $\forall w \in i (\llbracket |\varphi| \rrbracket^{w,i} = T)$, so it suffices to show that $\llbracket |\varphi^O| \rrbracket^{w,i} = T$ if $\llbracket |\varphi_1^I| \rrbracket^{w,i} = \dots = \llbracket |\varphi_n^I| \rrbracket^{w,i} = T$.

Here are a few cases:

$\neg\mathbf{Elim}_{\triangleright}$: $\llbracket i + P \triangleright \neg\neg Q \rrbracket^{w,i} = T$ iff $\forall w' \in i + P (\llbracket \neg\neg Q \rrbracket^{w',i+P} = T)$ iff

$\forall w' \in i + P (\llbracket Q \rrbracket^{w',i+P} = T) \text{ iff } \llbracket i + P \triangleright Q \rrbracket^{w,i} = T.$

\vee **Intro** $_{\triangleright}$: $\llbracket i + P \triangleright Q \rrbracket^{w,i} = T \text{ iff } \forall w' \in i + P (\llbracket Q \rrbracket^{w',i+P} = T) \text{ only if } \forall w' \in i + P (\llbracket Q \vee R \rrbracket^{w',i+P} = T) \text{ iff } \llbracket i + P \triangleright Q \vee R \rrbracket^{w,i} = T.$

\perp **Intro** $_{\triangleright}$: $\llbracket i + P \triangleright Q \rrbracket^{w,i} = \llbracket i + P \triangleright \neg Q \rrbracket^{w,i} = T \text{ iff } i + P = \emptyset \text{ iff } i + P \triangleright \perp \text{ iff } \llbracket i + P \triangleright \perp \rrbracket^{w,i} = T.$

\diamond **Elim**: $\llbracket \neg i \triangleright \neg P \rrbracket^{w,i} = T \text{ iff } \exists w' \in i (\llbracket P \rrbracket^{w',i} = T) \text{ iff } i + P \neq \emptyset.$ So $\llbracket \neg i \triangleright \neg P \rrbracket^{w,i} = \llbracket i + P \triangleright Q \rrbracket^{w,i} = T \text{ iff } \forall w' \in i + P (\llbracket Q \rrbracket^{w',i+P} = T) \text{ only if (since } i + P \neq \emptyset) \exists w' \in i (\llbracket \neg Q \rrbracket^{w',i} = F) \text{ iff } \llbracket \neg i \triangleright \neg Q \rrbracket^{w,i} = T.$

\wedge **Intro**: $\llbracket |\varphi| \rrbracket^{w,i} = \llbracket |\psi| \rrbracket^{w,i} = T \text{ iff } \llbracket |\varphi| \wedge |\psi| \rrbracket^{w,i} = T.$

\perp **Elim**: $i \triangleright \varphi_1^I \text{ iff } \forall w \in i (\llbracket i \triangleright \perp \rrbracket^{w,i} = T) \text{ iff } i = \emptyset \text{ only if } i \triangleright \varphi^O.$

The remaining cases are similar.

Next consider **Info**'s complex inferential rules involving subproofs. Since these rules license transitions from facts of the form $|\varphi^{s_i}| \vdash_{s_i}^* |\psi^{s_i}|$ or $|\varphi^{s_i}| \vdash_{s_i} |\psi^{s_i}|$ and (in some cases) input sentences $|\varphi_1^I|, \dots, |\varphi_n^I|$ to output sentence $|\varphi^O|$, it must be shown that $i \triangleright \varphi^O$ if $\Gamma \cup \{\varphi^{s_i}\} \models_I \psi^{s_i}$ for each s_i (where Γ is a set of sentences in $S_{\mathcal{L}_1}$ incorporated by i) and $i \triangleright \varphi_1^I \wedge \dots \wedge \varphi_n^I$. For \neg **Intro** $_{\triangleright}$, \vee **Elim** $_{\triangleright}$, and \Rightarrow **Intro**, Γ is a set of persistent sentences. Since $\emptyset \triangleright \varphi^O$, it thus suffices to show that $\llbracket |\varphi^O| \rrbracket^{w,i} = T$ when both $\llbracket |\varphi^{s_i}| \rrbracket^{w,i'} = T$ only if $\llbracket |\psi^{s_i}| \rrbracket^{w,i'} = T$ for each s_i and non-empty $i' \subseteq i$, and $\llbracket |\varphi_1^I| \rrbracket^{w,i} = \dots = \llbracket |\varphi_n^I| \rrbracket^{w,i} = T$.

\neg **Intro** $_{\triangleright}$: $i + P = \emptyset$ only if $i + P \triangleright \neg Q$ iff $\llbracket i + P \triangleright \neg Q \rrbracket^{w,i} = 1$, so assume $i + P \neq \emptyset$. Consider arbitrary $w^* \in i + P$ and let $i^* = \{w^*\}$. $\llbracket Q \rrbracket^{w^*,i+P} = T \text{ iff } i^* \triangleright Q \text{ iff } i^* + P \triangleright Q \text{ iff } \llbracket i + P \triangleright Q \rrbracket^{w,i^*} = T \text{ only if } \llbracket i + P \triangleright \perp \rrbracket^{w,i^*} = T \text{ iff } i^* \triangleright \perp.$ Hence $\llbracket Q \rrbracket^{w^*,i+P} = F$, and since w^* was arbitrary, $\forall w' \in i + P (\llbracket \neg Q \rrbracket^{w',i+P} = T)$, so $\llbracket i + P \triangleright \neg Q \rrbracket^{w,i} = T$.

\vee **Elim** $_{\triangleright}$: $i + P = \emptyset$ only if $\llbracket i + P \triangleright S \rrbracket^{w,i} = T$, so assume $i + P \neq \emptyset$. Consider $w^* \in i + P$ and let $i^* = \{w^*\}$. $\llbracket i + P \triangleright Q \vee R \rrbracket^{w,i} = T \text{ iff } i + P \triangleright Q \vee R \text{ only if } (i^* + P \triangleright Q \text{ or } i^* + P \triangleright R) \text{ iff } (\llbracket i + P \triangleright Q \rrbracket^{w,i^*} = T \text{ or } \llbracket i + P \triangleright R \rrbracket^{w,i^*} = T) \text{ only if } \llbracket i + P \triangleright S \rrbracket^{w,i^*} = T \text{ iff } i^* \triangleright S \text{ iff } \llbracket S \rrbracket^{w^*,i+P} = T.$ Since w^* was arbitrary, $\forall w' \in i + P (\llbracket S \rrbracket^{w',i+P} = T)$, so $\llbracket i + P \triangleright S \rrbracket^{w,i} = T$.

\Rightarrow **Intro**: $i + P = \emptyset$ only if $\llbracket i + P \triangleright Q \rrbracket^{w,i} = T$, so assume $i + P \neq \emptyset$. $\llbracket i \triangleright P \rrbracket^{w,i+P} = T$, so $\llbracket i \triangleright Q \rrbracket^{w,i+P} = T$, and so $\llbracket i + P \triangleright Q \rrbracket^{w,i} = T$.

For \neg **Intro** and \forall **Elim**, Γ can include non-persistent sentences. It then suffices to show that $\llbracket |\varphi^O| \rrbracket^{w,i} = T$ when both $\llbracket |\varphi^{s_i}| \rrbracket^{w,i} = T$ only if $\llbracket |\psi^{s_i}| \rrbracket^{w,i} = T$ for each s_i , and $\llbracket |\varphi_1^I| \rrbracket^{w,i} = \dots = \llbracket |\varphi_n^I| \rrbracket^{w,i} = T$.

\neg **Intro**: $\llbracket |\varphi| \rrbracket^{w,i} = T$ only if $\llbracket |i \triangleright \perp| \rrbracket^{w,i} = T$ iff $i \triangleright \perp$ iff $i = \emptyset$, contradicting the working assumption that $i \neq \emptyset$. Hence $\llbracket |\neg\varphi| \rrbracket^{w,i} = T$.

\forall **Elim**: $\llbracket |\varphi| \vee |\psi| \rrbracket^{w,i} = T$ iff either $\llbracket |\varphi| \rrbracket^{w,i} = T$ or $\llbracket |\psi| \rrbracket^{w,i} = T$. $\llbracket |\varphi| \rrbracket^{w,i} = T$ only if $\llbracket |\chi| \rrbracket^{w,i} = T$. $\llbracket |\psi| \rrbracket^{w,i} = T$ only if $\llbracket |\chi| \rrbracket^{w,i} = T$. Hence $\llbracket |\chi| \rrbracket^{w,i} = T$. \square

B.4 Completeness

I prove the following completeness theorem:

Thm 2. If $\{\varphi_1, \dots, \varphi_n\} \models_I \psi$, then $\{|\varphi_1|, \dots, |\varphi_n|\} \vdash_{\text{Info}} |\psi|$.

Proof: Assume $\{|\varphi_1|, \dots, |\varphi_n|\} \not\vdash_{\text{Info}} |\psi|$. I show $\{\varphi_1, \dots, \varphi_n\} \not\models_I \psi$. Where **Sent** is a standard proof system for sentential logic (such as the Fitch-style proof system \mathcal{F}_T in Barwise and Etchemendy [1999] minus the rules for \triangleright and \equiv), I follow the strategy of van der Does, Groeneveld, and Veltman [1997] and leverage the completeness of **Sent** to prove the general theorem.⁷

First consider the simple case where $\varphi_1, \dots, \varphi_n, \psi \in S_{\mathcal{L}_0}$. Given **Info**'s inner logic rules, $\{|\varphi_1|, \dots, |\varphi_n|\} \not\vdash_{\text{Info}} |\psi|$ only if $\{\varphi_1, \dots, \varphi_n\} \not\vdash_{\text{Sent}} \psi$ iff $\{w^*\} \triangleright \varphi_1 \wedge \dots \wedge \varphi_n \wedge \neg\psi$ for some w^* in logical space \mathbb{W} .

Next consider the case where $\varphi_1, \dots, \varphi_n, \psi$ are either sentences in $S_{\mathcal{L}_0}$ or are of the form $\Box P$, $\Diamond P$, or $P \Rightarrow Q$, where $P, Q \in S_{\mathcal{L}_0}$. We can define the following two functions:

$$\begin{array}{ll} P^\bullet & = P & P^{\bullet\bullet} & = P \\ \Box P^\bullet & = P & \Box P^{\bullet\bullet} & = P \\ \Diamond P^\bullet & = \neg\perp & \Diamond P^{\bullet\bullet} & = P \\ P \Rightarrow Q^\bullet & = \neg P \vee Q & P \Rightarrow Q^{\bullet\bullet} & = \neg P \vee Q \end{array}$$

$\{|\varphi|\} \vdash_{\text{Info}} |\varphi^\bullet|$ and $\{|\varphi^{\bullet\bullet}|\} \vdash_{\text{Info}} |\varphi|$. Hence $\{|\varphi_1|, \dots, |\varphi_n|\} \not\vdash_{\text{Info}} |\psi|$ only if $\{|\varphi_1^\bullet|, \dots, |\varphi_n^\bullet|\} \not\vdash_{\text{Info}} |\psi^{\bullet\bullet}|$ only if $\{\varphi_1^\bullet, \dots, \varphi_n^\bullet\} \not\vdash_{\text{Sent}} |\psi^{\bullet\bullet}|$ iff $\llbracket |\varphi_1^\bullet| \rrbracket^{w^*, \{w^*\}} = \dots = \llbracket |\varphi_n^\bullet| \rrbracket^{w^*, \{w^*\}} = T$ and $\llbracket |\psi^{\bullet\bullet}| \rrbracket^{w^*, \{w^*\}} = F$ for some $w^* \in \mathbb{W}$. Let $i_1^* = \{w \in \mathbb{W} : \llbracket |\varphi_1^\bullet| \rrbracket^{w, \{w\}} = \dots = \llbracket |\varphi_n^\bullet| \rrbracket^{w, \{w\}} = T\}$ and $i_2^* = i_1^* \cap \{w \in \mathbb{W} : \llbracket |\psi^{\bullet\bullet}| \rrbracket^{w, \{w\}} = F\}$. Note that $w^* \in i_2^* \subseteq i_1^*$.

⁷van der Does, Groeneveld, and Veltman work with only the weak language $\mathcal{L}_{0.5}$ with the following syntax:

Syntax of $\mathcal{L}_{0.5}$. The symbols of $\mathcal{L}_{0.5}$ are $A, B, C, \dots, \neg, \wedge, \diamond$, and parentheses. If P is a sentence in $S_{\mathcal{L}_0}$ not involving \vee and \perp , then $P, \diamond P \in S_{\mathcal{L}_{0.5}}$. Nothing else is in $S_{\mathcal{L}_{0.5}}$.

i_1^* and i_2^* are the information states we are after. Suppose ψ is *not* of the form $\diamond P$:

—If φ_i is of the form P or $\square P$, $\forall w \in i_1^*(\llbracket P \rrbracket^{w, i_1^*} = T)$, so $i_1^* \triangleright \varphi_i$.

—If φ_i is of the form $P \Rightarrow Q$, $\forall w \in i_1^*(\llbracket \neg P \vee Q \rrbracket^{w, i_1^*} = T)$, so $i_1^* \triangleright \varphi_i$.

—If φ_i is of the form $\diamond P$ and $i_1^* \not\triangleright \varphi_i$, then $\forall w \in i_1^*(\llbracket P \rrbracket^{w, i_1^*} = F)$, so $\{\varphi_1^\bullet, \dots, \varphi_n^\bullet\} \vdash_{\text{Sent}} \neg P$ by completeness. But $\{\varphi_1^\bullet, \dots, \varphi_n^\bullet\} \vdash_{\text{Sent}} \neg P$ only if $\{|\varphi_1^\bullet|, \dots, |\varphi_n^\bullet|\} \vdash_{\text{Info}} |\neg P|$ only if $\{|\varphi_1|, \dots, |\varphi_n|\} \vdash_{\text{Info}} |\neg P|$ only if $\{|\varphi_1|, \dots, |\varphi_n|\} \vdash_{\text{Info}} |\psi|$ (note $\{|\diamond P|, |\neg P|\} \vdash_{\text{Info}} \mathbf{i} \triangleright \perp$). Since this contradicts the assumption that $\{|\varphi_1|, \dots, |\varphi_n|\} \not\vdash_{\text{Info}} |\psi|$, $i_1^* \triangleright \varphi_i$.

—If ψ is of the form P or $\square P$, $\llbracket P \rrbracket^{w^*, i_1^*} = F$, so $i_1^* \not\triangleright \psi$.

—If ψ is of the form $P \Rightarrow Q$, $\llbracket P \rrbracket^{w^*, i_1^*} = T$ and $\llbracket Q \rrbracket^{w^*, i_1^*} = F$, so $i_1^* \not\triangleright \psi$.

Thus, $i_1^* \triangleright \varphi_1 \wedge \dots \wedge \varphi_n$, and $i_1^* \not\triangleright \psi$.

Next suppose ψ is of the form $\diamond P$:

—If φ_i is of the form P or $\square P$, $\forall w \in i_2^*(\llbracket P \rrbracket^{w, i_2^*} = T)$, so $i_2^* \triangleright \varphi_i$.

—If φ_i is of the form $P \Rightarrow Q$, $\forall w \in i_2^*(\llbracket \neg P \vee Q \rrbracket^{w, i_2^*} = T)$, so $i_2^* \triangleright \varphi_i$.

—If φ_i is of the form $\diamond P$ and $i_2^* \not\triangleright \varphi_i$, then $\forall w \in i_2^*(\llbracket P \rrbracket^{w, i_2^*} = F)$, so $\{\varphi_1^\bullet, \dots, \varphi_n^\bullet, \neg \psi^{\bullet\bullet}\} \vdash_{\text{Sent}} \neg P$. But $\{\varphi_1^\bullet, \dots, \varphi_n^\bullet, \neg \psi^{\bullet\bullet}\} \vdash_{\text{Sent}} \neg P$ only if $\{|\varphi_1^\bullet|, \dots, |\varphi_n^\bullet|, |P|\} \vdash_{\text{Info}} |\psi^{\bullet\bullet}|$ only if $\{|\varphi_1|, \dots, |\varphi_n|\} \vdash_{\text{Info}} |P \Rightarrow \psi^{\bullet\bullet}|$ only if $\{|\varphi_1|, \dots, |\varphi_n|\} \vdash_{\text{Info}} |\diamond \psi^{\bullet\bullet}|$ given $\diamond \mathbf{Elim}$. Since this contradicts the assumption that $\{|\varphi_1|, \dots, |\varphi_n|\} \not\vdash_{\text{Info}} |\psi|$, $i_2^* \triangleright \varphi_i$.

—Since ψ is of the form $\diamond P$, $\forall w \in i_2^*(\llbracket P \rrbracket^{w, i_2^*} = F)$, so $i_2^* \not\triangleright \psi$.

Thus, $i_2^* \triangleright \varphi_1 \wedge \dots \wedge \varphi_n$, and $i_2^* \not\triangleright \psi$.

Finally consider the general case where $\varphi_1, \dots, \varphi_n, \psi \in S_{\mathcal{L}_1}$. Note the following facts, where $\Gamma \subseteq S_{\mathcal{L}_1}$ and $|\Gamma| \subseteq S_{\mathcal{L}_2}$ is obtained from Γ by applying $|\cdot|$ to each of its members:

—If $\chi_1 \notin S_{\mathcal{L}_0}$, then $|\Gamma| \not\vdash_{\text{Info}} |\chi_1|$ only if $|\Gamma| \cup \{|\neg \chi_1|\} \not\vdash_{\text{Info}} |\perp|$. Also, $i \triangleright \neg \chi_1$ iff $i \not\triangleright \chi_1$ when $i \neq \emptyset$, since $\llbracket |\neg \chi_1| \rrbracket^{w, i} = T$ iff $\llbracket |\neg \chi_1| \rrbracket^{w, i} = T$ iff $\llbracket |\chi_1| \rrbracket^{w, i} = F$. Hence $\Gamma \cup \{\neg \chi_1\} \not\vdash_I \perp$ only if $\Gamma \not\vdash_I \chi_1$.

—If χ_1 is not of the form P , $\square P$, $\diamond P$, or $P \Rightarrow Q$, $|\Gamma| \cup \{|\chi_1|\} \not\vdash_{\text{Info}} |\chi_2|$ only if there is a sentence χ'_1 which is a disjunction of conjunctions of sentences or negated sentences of the form $\square P$, $\diamond P$, or $P \Rightarrow Q$, and $|\Gamma| \cup \{|\chi'_1|\} \not\vdash_{\text{Info}} |\chi_2|$. Also, $i \triangleright \chi_1$ iff $i \triangleright \chi'_1$, since $\llbracket |\chi_1| \rrbracket^{w, i} = \llbracket |\chi'_1| \rrbracket^{w, i}$. Hence $\Gamma \cup \{\chi'_1\} \not\vdash_I \chi_2$ only if $\Gamma \cup \{\chi_1\} \not\vdash_I \chi_2$.

—If $\chi_1 \notin S_{\mathcal{L}_0}$ and $\chi_2 \notin S_{\mathcal{L}_0}$, then $|\Gamma| \cup \{|\chi_1 \vee \chi_2|\} \not\vdash_{\text{Info}} |\chi_3|$ only if either $|\Gamma| \cup \{|\chi_1|\} \not\vdash_{\text{Info}} |\chi_3|$ or $|\Gamma| \cup \{|\chi_2|\} \not\vdash_{\text{Info}} |\chi_3|$. Also, $i \triangleright \chi_1 \vee \chi_2$ iff either $i \triangleright \chi_1$ or $i \triangleright \chi_2$, since $\llbracket |\chi_1 \vee \chi_2| \rrbracket^{w, i} = T$ iff $\llbracket |\chi_1| \vee |\chi_2| \rrbracket^{w, i} = T$

iff either $\llbracket |\chi_1| \rrbracket^{w,i} = T$ or $\llbracket |\chi_2| \rrbracket^{w,i} = T$. Hence $\Gamma \cup \{\chi_1\} \not\models_I \chi_3$ only if $\Gamma \cup \{\chi_1 \vee \chi_2\} \not\models_I \chi_3$, and $\Gamma \cup \{\chi_2\} \not\models_I \chi_3$ only if $\Gamma \cup \{\chi_1 \vee \chi_2\} \not\models_I \chi_3$.

—If $\chi_1 \notin S_{\mathcal{L}_0}$ and $\chi_2 \notin S_{\mathcal{L}_0}$, then $|\Gamma| \cup \{|\chi_1 \wedge \chi_2|\} \not\models_{\text{Info}} |\chi_3|$ only if $|\Gamma| \cup \{|\chi_1|, |\chi_2|\} \not\models_{\text{Info}} |\chi_3|$. Also, $i \triangleright \chi_1 \wedge \chi_2$ iff $i \triangleright \chi_1$ and $i \triangleright \chi_2$, since $\llbracket |\chi_1 \wedge \chi_2| \rrbracket^{w,i} = T$ iff $\llbracket |\chi_1| \wedge |\chi_2| \rrbracket^{w,i} = T$ iff $\llbracket |\chi_1| \rrbracket^{w,i} = \llbracket |\chi_2| \rrbracket^{w,i} = T$. Hence $\Gamma \cup \{\chi_1, \chi_2\} \not\models_I \chi_3$ only if $\Gamma \cup \{\chi_1 \wedge \chi_2\} \not\models_I \chi_3$.

— $|\Gamma| \cup \{|\neg \Box P|\} \not\models_{\text{Info}} |\chi_1|$ only if $|\Gamma| \cup \{|\Diamond \neg P|\} \not\models_{\text{Info}} |\chi_1|$. Also, $i \triangleright \neg \Box P$ iff $i \triangleright \Diamond \neg P$, so $\Gamma \cup \{\Diamond \neg P\} \not\models_I \chi_1$ only if $\Gamma \cup \{\neg \Box P\} \not\models_I \chi_1$.

— $|\Gamma| \cup \{|\neg \Diamond P|\} \not\models_{\text{Info}} |\chi_1|$ only if $|\Gamma| \cup \{|\Box \neg P|\} \not\models_{\text{Info}} |\chi_1|$. Also, $i \triangleright \neg \Diamond P$ iff $i \triangleright \Box \neg P$, so $\Gamma \cup \{\Box \neg P\} \not\models_I \chi_1$ only if $\Gamma \cup \{\neg \Diamond P\} \not\models_I \chi_1$.

— $|\Gamma| \cup \{|\neg(P \Rightarrow Q)|\} \not\models_{\text{Info}} |\chi_1|$ only if $|\Gamma| \cup \{|\Diamond(P \wedge \neg Q)|\} \not\models_{\text{Info}} |\chi_1|$. Also, $i \triangleright \neg(P \Rightarrow Q)$ iff $i \triangleright \Diamond(P \wedge \neg Q)$, so $\Gamma \cup \{\Diamond(P \wedge \neg Q)\} \not\models_I \chi_1$ only if $\Gamma \cup \{\neg(P \Rightarrow Q)\} \not\models_I \chi_1$.

Therefore, there are sentences χ_i^j of the form P , $\Box P$, $\Diamond P$, or $P \Rightarrow Q$ such that $\{|\varphi_1|, \dots, |\varphi_n|\} \not\models_{\text{Info}} |\psi|$ only if $\bigvee_j (\{|\chi_1^j|, \dots, |\chi_{n_j}^j|\} \not\models_{\text{Info}} |\chi_{n_j+1}^j|)$, and $\{\chi_1^j, \dots, \chi_{n_j}^j\} \not\models_I \chi_{n_j+1}^j$ only if $\{\varphi_1, \dots, \varphi_n\} \not\models_I \psi$ for all j . From the completeness result proven above, it follows that $\{\varphi_1, \dots, \varphi_n\} \not\models_I \psi$. \square

Bibliography

- Ernest W. Adams. *The Logic of Conditionals*. Reidel, Dordrecht, 1975.
- Alan Ross Anderson, Nuel Belnap Jr., and J. Michael Dunn. *Entailment, Volume II*. Princeton University Press, Princeton, 1992.
- David Barnett. Ramsey + Moore \neq God. *Analysis*, 68(2):168–174, 2008.
- Jon Barwise and John Etchemendy. *Language, Proof and Logic*. CSLI Publications, Stanford, 1999.
- Jonathan Bennett. *A Philosophical Guide to Conditionals*. Clarendon Press, Oxford, 2003.
- Richard Bradley. A defence of the Ramsey test. *Mind*, 116(461):1–21, 2007.
- John Broome. Normative requirements. *Ratio*, 12:398–419, 1999.
- Lewis Carroll. A logical paradox. *Mind*, 3(11):436–438, 1894.
- Lewis Carroll. What the tortoise said to Achilles. *Mind*, 4(14):278–280, 1895.
- David Christensen. *Putting Logic in its Place: Formal Constraints on Rational Belief*. Oxford University Press, Oxford, 2004.
- Ivano Ciardelli and Floris Roelofson. Inquisitive logic. *Journal of Philosophical Logic*, 40:55–94, 2011.
- Donald Davidson. Incoherence and irrationality. In *Problems of Rationality*, pages 189–198. Oxford University Press, 2004.
- Keith DeRose. Epistemic possibilities. *Philosophical Review*, 100(4): 581–605, 1991.
- Dorothy Edgington. Do conditionals have truth conditions? *Critica: Revista Hispanoamericana de Filosofia*, 18(52):3–39, 1986.

- Andy Egan. Epistemic modals, relativism, and assertion. *Philosophical Studies*, 133(1):1–22, 2007.
- Jonathan St. B. T. Evans and D. E. Over. *If: Supposition, Pragmatics, and Dual Processes*. Oxford University Press, Oxford, 2004.
- Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning About Knowledge*. MIT Press, Cambridge, 1995.
- Hartry Field. Truth and the unprovability of consistency. *Mind*, 115 (459):567–606, 2006.
- Hartry Field. *Saving Truth From Paradox*. Oxford University Press, Oxford, 2008.
- Hartry Field. What is the normative role of logic? *Proceedings of the Aristotelian Society Supplementary Volume*, LXXXIII:251–268, 2009a.
- Hartry Field. Pluralism in logic. *Review of Symbolic Logic*, 2:342–359, 2009b.
- Hartry Field. What is logical validity? unpublished manuscript.
- Richard Foley. Beliefs, degrees of belief, and the Lockean thesis. In Franz Huber and Christoph Schmidt-Petri, editors, *Degrees of Belief*, pages 37–47. Springer, 2009.
- Gottlob Frege. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. L. Nebert, Halle, 1879. Partially translated by Peter Geach in *Translations from the Philosophical Writings of Gottlob Frege*, pages 1–20, 1960.
- Gottlob Frege. Logik. 1879-1891. Translated as ‘Logic’ in *Posthumous Writings*, pages 1–8, 1979.
- Gottlob Frege. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*. Pohle, Jena, 1893. Partially translated as *The Basic Laws of Arithmetic: Exposition of the System*, 1964.
- Gottlob Frege. Logik. 1897. Translated as ‘Logic’ in *Posthumous Writings*, pages 126–151, 1979.
- Gottlob Frege. Der Gedanke: Eine logische Untersuchung. *Beiträge zur Philosophie des Deutschen Idealismus*, 1, 1918. Translated as ‘Thoughts’ by Peter Geach and R. H. Stoothoff in *Collected Papers on Mathematics, Logic, and Philosophy*, pages 351–372, 1984.

- Gottlob Frege. *Translations from the Writings of Gottlob Frege*. Eds. Peter Geach and Max Black, Blackwell, Oxford, 1960.
- Gottlob Frege. *The Basic Laws of Arithmetic: Exposition of the System*. Trans. Montgomery Furth, University of California Press, Berkeley, 1964.
- Gottlob Frege. *Posthumous Writings*. Trans. Peter Long and Roger White, University of Chicago Press, Chicago, 1979.
- Gottlob Frege. *Collected Papers on Mathematics, Logic, and Philosophy*. Ed. Brian McGuinness, Blackwell, Oxford, 1984.
- Allan Gibbard. *Wise Choices, Apt Feelings*. Harvard University Press, Cambridge, 1990.
- Jeroen Groenendijk. The logic of interrogation. In Tanya Matthews and Devon Strolovitch, editors, *Proceedings of SALT IX*, pages 109–126. CLC Publications, 1999.
- Jeroen Groenendijk and Martin Stokhof. *Studies on the semantics of questions and the pragmatics of answers*. PhD thesis, University of Amsterdam, 1984.
- Ian Hacking. Possibility. *Philosophical Review*, 76(2):143–168, 1967.
- C. L. Hamblin. Questions. *Australasian Journal of Philosophy*, 36(3): 159–168, 1958.
- Gilbert Harman. *Change in View: Principles of Reasoning*. MIT Press, Cambridge, 1986.
- Wesley Holliday and Thomas Icard III. Measure semantics and qualitative semantics for epistemic modals. unpublished manuscript.
- Joris Hulstijn. Structured information states—raising and resolving issues. In Anton Benz and Gerhard Jäger, editors, *Proceedings of MunDial'97*, pages 99–117. 1997.
- James Isaacs and Kyle Rawlins. Conditional questions. *Journal of Semantics*, 25(3):269–319, 2008.
- James M. Joyce. *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge, 1999.

- David Kaplan. Demonstratives. In Joseph Almog, John Perry, and Howard Wettstein, editors, *Themes From Kaplan*, pages 481–563. Oxford University Press, 1989.
- Niko Kolodny. Why be rational? *Mind*, 114(455):509–563, 2005.
- Niko Kolodny. How does coherence matter? *Proceedings of the Aristotelian Society*, CVII(3):229–263, 2007.
- Niko Kolodny. Why be disposed to be coherent? *Ethics*, 118(3):437–463, 2008.
- Niko Kolodny and John MacFarlane. Ifs and oughts. *Journal of Philosophy*, CVII(3):115–143, 2010.
- Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72(19):690–716, 1975.
- Hannes Leitgeb. God - Moore = Ramsey (a reply to Chalmers and Hájek). *Mind*, 118(469):71–100, 2011.
- David Lewis. *Counterfactuals*. Harvard University Press, Cambridge, 1973.
- David Lewis. Adverbs of quantification. In E. L. Keenan, editor, *Semantics of Natural Language*, pages 3–15. Cambridge University Press, 1975.
- David Lewis. Index, context, and content. In Stig Kanger and Sven Öhman, editors, *Philosophy and Grammar*, pages 79–100. Reidel, 1980.
- David Lewis. Logic for equivocators. *Noûs*, 16(3):431–441, 1982.
- John MacFarlane. Frege, Kant, and the logic in Logicism. *Philosophical Review*, 111(1):25–65, 2002.
- John MacFarlane. Future contingents and relative truth. *Philosophical Quarterly*, 53(212):321–336, 2003.
- John MacFarlane. Epistemic modals are assessment-sensitive. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*, pages 144–178. Oxford University Press, 2011.

- John MacFarlane. Logical constants. In Edward N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Fall 2009. <http://plato.stanford.edu/archives/fall2009/entries/logical-constants/>.
- John MacFarlane. In what sense (if any) is logic normative for thought? unpublished manuscript.
- John L. Mackie. *Truth, Probability, and Paradox*. Clarendon Press, Oxford, 1973.
- Vann McGee. A counterexample to modus ponens. *Journal of Philosophy*, 82(9):462–471, 1985.
- Frank Ramsey. General propositions and causality. 1929. In D. H. Mellor, editor, *Philosophical Papers*, pages 145–163. Cambridge University Press, 1990.
- Nishi Shah. How truth governs belief. *Philosophical Review*, 112(4):447–482, 2003.
- Robert Stalnaker. A theory of conditionals. In Nicholas Rescher, editor, *Studies in Logical Theory*, pages 98–112. Oxford University Press, 1968.
- Robert Stalnaker. Indicative conditionals. *Philosophia*, 5(3):269–286, 1975.
- Robert Stalnaker. Assertion. In P. Cole, editor, *Syntax and Semantics 9: Pragmatics*, pages 315–332. Academic Press, 1978.
- Robert Stalnaker. Indexical belief. *Synthese*, 49(1):129–151, 1981.
- Robert Stalnaker. *Inquiry*. MIT Press, Cambridge, 1984.
- Robert Stalnaker. Common ground. *Linguistics and Philosophy*, 25:701–721, 2002.
- Alfred Tarski. O pojęciu wynikania logicznego. *Przegląd Filozoficzny*, 39, 1936a. Translated as ‘On the Concept of Logical Consequence’ in *Logic, Semantics, Metamathematics*, pp. 409–420, 1956.
- Alfred Tarski. Der wahrheitsbegriff in den formalisierten sprachen. *Studia Philosophica*, 1, 1936b. Translated as ‘The Concept of Truth in Formalized Languages’ in *Logic, Semantics, Metamathematics*, pp. 152–278, 1956.

- Alfred Tarski. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938 by Alfred Tarski*. Ed. and Trans. J. H. Woodger, Clarendon Press, Oxford, 1956.
- Paul Teller. Epistemic possibility. *Philosophia*, 2(4):303–320, 1972.
- Jan Łukasiewicz. Philosophische bemerkungen zu mehrwertigen systemen des aussagenkalküls. *Comptes rendus des seances de la Societe des Sciences et des Lettre de Varsovie*, 23, 1930. Translated as ‘Philosophical Remarks on Many-Valued Systems of Propositional Logic’ in *Selected Works*, pp. 153–178, 1970.
- Jan Łukasiewicz. *Selected Works*. North Holland, Amsterdam, 1970.
- Jaap van der Does, Willem Groeneveld, and Frank Veltman. An update on ‘might’. *Journal of Logic, Language, and Information*, 6:361–380, 1997.
- Bas van Fraassen. Review of Brian Ellis, *Rational Belief Systems*. *Canadian Journal of Philosophy*, 10(3):497–511, 1980.
- Sophia Velissaratou. Conditional questions and which-interrogatives. Master’s thesis, University of Amsterdam, 2000.
- Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.
- Kai von Fintel and Anthony S. Gillies. ‘Might’ made right. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*, pages 108–130. Oxford University Press, 2011.
- Roger White. Epistemic permissiveness. *Philosophical Perspectives*, 19: 445–459, 2005.
- Malte Willer. New surprises for the Ramsey test. *Synthese*, 176(2): 291–309, 2010.
- Malte Willer. A remark on iffy oughts. *Journal of Philosophy*, 109(7): 449–461, 2012.
- Malte Willer. Dynamics of epistemic modality. *Philosophical Review*, 122(1):45–92, 2013.
- Timothy Williamson. Knowing and asserting. *Philosophical Review*, 105 (4):489–523, 1996.

- Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, Oxford, 2000.
- Crispin Wright. Facts and certainty. *Proceedings of the British Academy*, 71:429–472, 1985.
- Seth Yalcin. Epistemic modals. *Mind*, 116(464):983–1026, 2007.
- Seth Yalcin. Probability operators. *Philosophy Compass*, 5(11):916–937, 2010.
- Seth Yalcin. Nonfactualism about epistemic modality. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*, pages 295–332. Oxford University Press, 2011.
- Seth Yalcin. A counterexample to modus tollens. *Journal of Philosophical Logic*, 41(6):1001–1024, 2012a.
- Seth Yalcin. Context probabilism. In Maria Aloni, Vadim Kimmelman, Floris Roelofsen, Galit Weidman Sassoon, Katrin Schulz, and Matthijs Westera, editors, *Proceedings of the Eighteenth Amsterdam Colloquium*, volume 7218 of *Lecture Notes in Computer Science*, pages 12–21. Springer, 2012b.
- Seth Yalcin. Bayesian expressivism. *Proceedings of the Aristotelian Society*, CXII(2):123–160, 2012c.