1 Why Entailment?

1.1 Planning Proofs

I want to prove that a statement B holds in a particular theory. I think I can show that A follows from the axioms of that theory and I think that I can prove that A entails B. In order to prove B, I formulate a **proof plan**: *If A entails B, then if Ax (the conjoined axioms of the theory) entails A, then Ax entails B*. The 'if ... then's in this statement of the proof plan themselves do not express a standard indicative conditional. Rather, they express logical entailments. Formalising logical entailment with an arrow, $\rightarrow$, this proof plan can be written

$$(A \rightarrow B) \rightarrow ((Ax \rightarrow A) \rightarrow (Ax \rightarrow B)).$$

What makes this a good proof plan? Part of the answer to this is that a proof plan is good only when it is a theorem of the logic of entailment. Proofs have to be logically valid arguments and the salient notion of validity is captured by the logic of entailment. This book is about this logic and its philosophical and formal semantics.

Of course, there are also non-logical criteria that go into determining whether a proof plan is good. As Saunders Mac Lane[1] said:

> A proof for a given theorem is not a haphazard collection of individual steps, taken arbitrarily one after another, as the classical logic might easily lead us to believe. On the contrary, there is some definite reason for the inclusion of each one of these steps in the proof; that is, each individual step is taken for some specific purpose. [99, p 125]

The rationale for the ordering of the steps in a proof may be logical. It may have to do with how they fit together logically into the proof. But it may have to do with psychological factors that govern how people understand proofs. The construction of logic of proof plans involves the abstraction away from these psychological factors to produce a theory of the formal requirements of a good proof.

One might object that we could attribute every feature that makes a proof good that is not available in classical logic, say, to the contingencies of human psychology and merely say that classical logic is the theory of proof plans. In this book, I suggest that classical logic and modal logics based on classical logic do not make certain distinctions or treat the connectives in ways that make a logic useful in distinguishing between good and bad proof plans.

[1] I take the terminology of proof plans from Mac Lane, who used the very similar "plans of proof".

Suppose that B really does follow from the axioms of the theory (Ax). Then, if we treat $\to$ as the classical conditional or as a classical form of strict implication (see Chapters 2 and 3), the following inference is valid:

$$\frac{\vdash Ax \to B}{\vdash C \to (Ax \to B)}$$

for arbitrary formulas, C. Treating this $\to$ as entailment means that we treat $C \to (Ax \to B)$ as a good proof plan, at least as far as the formal logic of the proof plan is concerned. The acceptance of this rule and its classical kin severely limit the usefulness of a logic of entailment. It is not the case that mathematicians or others would accept this $C \to (Ax \to B)$ as a good proof plan when C is unrelated to Ax or B. Thus, accepting classical logic or one of its modal extensions would force us to give an almost entirely psychological and pragmatic analysis of proof plans concerning logical theorems. Surely, a formal logic of entailment can contribute a lot more than this.

As the above indicates, in order to formulate proof plans, I use an *entailment connective*. I need to be able to nest expressions about proving one formula from another within an entailment. Thus, I have entailments nested within entailments. All logical systems give rise to one or more entailment relations. A *logic of entailment* contains an entailment connective.

In order to develop and justify a logic of entailment, I have to delve very deeply into the problem of nested entailments. In this book, I adopt G. E. Moore's definition of 'entailment' as the 'converse of deducibility' [147, p 291]. On this definition,

$$\text{`}A \to B\text{' is true if and only if } A \Vdash B.$$

This truth condition does not help when it comes to the meanings of nested entailments. For it does not make sense in our standard logical languages to write, say,

$$(Ax \Vdash A) \Vdash (Ax \Vdash B).$$

It is the central aim of this book to develop a proof theory and a semantic theory that both make sense of the claim that the entailment connective expresses a reasonable deducibility relation and characterises a logic of entailment that is robust and useful in formulating proof plans. The proof theory that I adopt is Alan Anderson and Nuel Belnap's natural deduction system for their logic E of relevant entailment [3, 4] together with a labelled natural deduction system which is closely tied to the semantics. The semantics that I employ is a version of Kit Fine's "Models for Entailment" [61] modified to incorporate Robert Goldblatt's and my semantics for quantification [134].

1.2 Logic and Intuition

When trying to create a formal logic of any concept we need some criteria of success and of failure. We need to be able to tell, at least within some boundaries, whether we have captured the concept that we want to represent. The most common benchmark used to justify or reject theories of entailment employed over the last century was whether this notion captured the intuitive notion of a deduction. Some philosophers have claimed, however, that our intuitions regarding deduction conflict with one another. One philosopher who said this was Casimir Lewy:

> [T]here is in any case no theory of entailment, however complex, which would enable us to accept all those entailment propositions which on an intuitive level we wish to accept and reject all those which on an intuitive level we wish to reject. And, so far as I can see, there is no *hope* of such a theory. [116, pp 133–134]

Timothy Smiley agreed with Lewy [194]. To support this view, Smiley gave the following list of theses about entailment, the first four are strongly intuitive, but the last is strongly counter-intuitive:

1. A entails $A \vee B$.
2. If A entails B, then $A \wedge C$ entails $B \wedge C$.
3. $(A \vee B) \wedge \neg A$ entails B.
4. If A entails B and B entails C, then A entails C.
5. $A \wedge \neg A$ entails B.

He said about this list that "if [we are] uncorrupted by acquaintance with the polemical literature", we should accept 1–4 all hold but reject 5. The problem, however, was that 5 followed from 1–4.

Smiley did try to construct a theory that almost captured our intuitions regarding entailment. It did reject one thesis – the fourth. His logic did not accept the transitivity of entailment.[2] I argue for the acceptance of transitivity of entailment in Section 1.11 and in Chapter 7. But I set aside my disagreement with Smiley for a moment to discuss a methodological issue. Lewy and Smiley are right that there are certain intuitive theses and rules of inference that cannot be included in a logic of entailment without that logic's thereby containing some counter-intuitive theses or rules. Among these counter-intuitive theses are the so-called paradoxes of strict implication that are discussed at length in this book.

Lewy and Smiley attempted to weigh intuitions against one another to justify the logics that they accept. Smiley argued for his non-transitive logic on this basis and Lewy argued for C. I. Lewis's theory of strict implication, saying that its counter-intuitive consequences were "less counter-intuitive than the consequences of accepting any of the alternative theories" [116, p 134].

I find, however, that this weighing of intuitions, taken alone, is very unsatisfying. It may be that in any justification of a logical system, one has to make use of some intuitions. The problem with the use of intuitions, however, is that they can be malleable or ephemeral. As Lewy pointed out [116, p 108], Lewis had originally rejected the theses that all propositions entailed every tautology and that every contradiction entailed every proposition, but after producing his so-called independent proofs (see

[2] Others have also rejected the transitivity of entailment for similar reasons, in particular Neil Tennant [204] and David Ripley [181].

https://doi.org/10.1017/9781009375283.002 Published online by Cambridge University Press

Chapter 2) he came to accept these supposed entailments. Lewis's argument was that the premises that he used in the independent proofs were so intuitive that they overruled his previous rejection of the paradoxes of strict implication. But as I also discuss in Chapter 2, Lewis's students Everett Nelson and William Parry rejected certain of Lewis's intuitions and created alternative logical systems that did not contain these paradoxical theses. This example illustrates the limitations of the method of playing off intuitions against one another. There seems to be no reason why someone else, who sees things slightly differently, will not have different weightings attributed to these various supposed logical principles and reject some of the premises of the independent arguments instead.

There are, however, other sorts of evidence that can be used. In this book, I appeal quite often to the jobs that I want a logic of entailment to do. If the logic is suited well to doing the tasks that are set out for it, this is good evidence for its acceptability.

1.3 Using Proof Plans to Justify the Choice of Logic

The project of representing proof plans is one of the central motivators for creating a theory of entailment. The notion of a proof, as something that is certain and settled, justifies the two principles that I discuss in Sections 1.9 and 1.10.

The first of these is the principle of inclusion, which says that if a proposition A is in a theory, even after inferences are made, A remains in that theory. The point of making inferences from, say, axioms of a theory is to discover the propositions other than the axioms that are in the theory. But these inferences do not change the fact that the axioms are in the theory. I discuss this feature of entailment at greater length in Section 1.9, but here I wish to discuss a methodological issue. This motivation for the principle of inclusion is not a mere appeal to intuitions about proof, but an appeal to the reason for proofs and the structure of proofs as they are commonly understood. We commonly understand proofs about theories as a way of increasing our understanding of proofs by discovering more propositions that are in those theories.

When we appeal to the tasks that a logic must be able to accomplish, rather than just intuitions, we have a better (or at least a second) measure of the viability of that logic. It may be that a logic cannot do everything that seems to be included in its job description, but we can certainly talk about what makes a logic overall the best at doing its job. With regard to the formulation of proof plans, a logic's having a good proof theory is one of those things that makes it good at being the logic of proof plans.

By 'proof theory', I do not mean merely a good set of axioms, all of which seem intuitive, but rather a structure, like a natural deduction system or a Gentzen sequent system, which gives a basic structure into which particular rules fit (or fail to fit). And this structure (and the rules) must be able to be explained in terms of possible proof plans. Consider again the plan to prove that $A \rightarrow B$ and $Ax \rightarrow B$ in order to show that $Ax \rightarrow B$. Here is a Fitch-style natural deduction proof:[3]

[3] For a brief introduction to natural deduction of this sort, see Chapter 4.

https://doi.org/10.1017/9781009375283.002 Published online by Cambridge University Press

1	$A \to B$	hypothesis
2	$Ax \to A$	hypothesis
3	Ax	hypothesis
4	$Ax \to A$	2, reiteration
5	A	3,4, $\to$E
6	$A \to B$	1, reiteration
7	B	5,6, $\to$E
8	$Ax \to B$	3–7, $\to$I
9	$(Ax \to A) \to (Ax \to B)$	2–8, $\to$I
10	$(A \to B) \to ((Ax \to A) \to (Ax \to B))$	1–9, $\to$I

The structure of this derivation both shows what needs further proof ($A \to B$ and $Ax \to A$) and what is assumed, as well as those that are to be used in a proof of B and, through its use of the entailment introduction and elimination rules it shows how the entailment connective represents the structure of the derivation. In this way, a good proof theory can show how an entailment connective can represent deducibility and show how the various principles of the logic are generated and show that they make up a coherent whole.

As I said above, the classical rule that makes any formula entail a theorem makes the logic of entailment of limited use. Thus, a proof theory that does not make this rule valid has, prima facie, an advantage over theories that do make it valid. It is for this reason that I adopt Anderson and Belnap's logic E and their natural deduction system for it. On a classically based natural deduction system, the following derivation might be valid:

1	$Ax \to B$	hypothesis
2	C	hypothesis
3	$Ax \to B$	1, reiteration
4	$C \to (Ax \to B)$	2–3, $\to$I
5	$(Ax \to B) \to (C \to (Ax \to C))$	1–4, $\to$I

If this is a theorem of the logic of entailment, then we have to accept the rule that $C \to (Ax \to B)$ is derivable from $Ax \to B$. The problem with this derivation is that the second hypothesis, C, is not used to derive $Ax \to B$. C is merely hypothesised and then in line 4 its hypothesis is discharged. It has not been used and therefore it has not earned the right to be discharged.

The mechanism by which Anderson and Belnap keep track of which hypotheses are really used in derivations is explained in detail in Chapter 4. That mechanism,

https://doi.org/10.1017/9781009375283.002 Published online by Cambridge University Press

however, requires further interpretation and explanation. This interpretation and explanation is given by the semantic theory that I present in Chapters 8–10.

1.4 Theory Closure

Another role for the logic of entailment is as a theory of theory closure. In this book, I claim that this function of the logic of entailment is not only one of its central functions, but understanding the relationship between entailment and theories is a key to understanding the meaning of the entailment connective itself.

Proofs and proof plans are often about theories. Logicians are often concerned with mathematical theories. Arithmetic and set theory are philosophers' favourites, but logicians have formalised other mathematical theories, such as algebraic and geometrical theories. The logical positivists and structuralist philosophers of science also had the goal to formalise the theories of physical and even social sciences.

I follow C. I. Lewis [106] and Richard Routley [182] in holding that theories, such as scientific theories, are closed under the principles of entailment. As I have said, logicians and others often prove things about the contents of theories. Among other things, I think that reflecting on the nature of theories allows us to find important restrictions to place on a theory of entailment. For example, in pre-1750s chemical or alchemical theories, there is no mention of oxygen or hydrogen. We cannot attribute to them the idea that 'lakes are filled with water' means the same thing as 'lakes are filled with H_2O', nor can we even attribute the latter sentence to theories of this period.

This example shows that theories cannot be understood as sets of propositions, at least not in the contemporary view of propositions. Moreover, it shows that a logic of entailment is not closed under replacement for identical propositions. This means that just because p and q happen to express the same proposition (in the contemporary sense), it may not be that the formula $p \leftrightarrow q$ is a theorem of the logic, where the double-ended arrow is co-entailment. I deal with this issue in Chapter 10 by introducing a metaphysically thin notion of proposition, as opposed to the metaphysically thick conception of proposition that is discussed in much of the contemporary literature.

I use a double turnstile, $\Vdash$, to represent the relation under which theories are closed. This is one of three consequence relations I associate with a logic. The turnstile, $\vDash$, is used to represent semantic consequence. And the single turnstile $\vdash$ is used to indicate theorems (as in '$\vdash A$', which means 'A is a theorem') and is such that '$A_1; \ldots; A_n \vdash B$' means '$\vdash A_1 \to (\ldots(A_n \to B)\ldots)$'. Where Γ is a set of formulas, $\Gamma \Vdash C$ if and only if there are some formulas $G_1, \ldots, G_n$ such that $G_1 \wedge \ldots \wedge G_n \vdash C$. In classically based logics (such as the standard modal logics) and intuitionist logic, $\vdash$ and $\Vdash$ are the same. But this is not true for relevant logics, as I show in Chapter 4.

For my project, it is crucial that the consequence relation (understood syntactically or semantically) is compact. That is, if C is derivable from a collection of premises Γ, then there is a finite sub-collection Γ', such that C is derivable from Γ'. In Chapters 6 and 10, I reply to arguments due to Alfred Tarski and Alonzo Church for the thesis that the consequence relation should not be compact. Their arguments are meant to show that any theory of theory closure must be essentially infinitary, that some inferences with infinitely many premises must be included in this theory that cannot be truncated in any way to finitary inferences.

I find the connection between theories and the logic of entailment to be particularly fruitful. I use the job of theory closure to indicate which properties a logic of entailment needs to have. And I use Tarski's concept of a *theory closure operator* to help provide a semantics of nested entailments. Each theory determines a consequence operator. For example, if T_1 is a theory that contains the formula $A \rightarrow B$, then the consequence operator C_1 is such that, if T_2 contains A, then B is in $C_1(T_2)$. The relationships between these consequence operators and theories can be seen to characterise a logical system. In Chapters 7 and 8, I show how natural conditions on theories and consequence operators can produce semantics for two relevant logics: one which I call *Generalised Tarski Logic* (GTL) and the other is Anderson and Belnap's logic E. Postulates governing the behaviour of consequence operators in relation to one another (and to themselves) give us different properties of nested entailments.

1.5 Entailment and Metaphysics

Like other theories, metaphysical theories are closed under true entailments. But there are other uses of the logic of entailment in metaphysics. The entailment connective has been used to formulate important theses in metaphysical theories. For example, G. E. Moore coined the term 'entailment' with its current philosophical meaning in order to formulate a theory of internal properties and relations. J. M. Dunn modified Moore's theory by using a relevant logic of entailment to develop a theory of "relevant predication" [3, 54, 55]. The concept of entailment has also been used in the contemporary theory of truthmakers [118, §1.1]. A truthmaker is something that makes a sentence or a proposition (i.e. a truth-bearer) true. Truthmakers have been said to entail the truth of their corresponding truth-bearers.

I don't think that any of these three uses are compatible with the use of entailment to capture the structure of proof or the structure of theories. I suggest, rather, that a different semantics should be used to understand entailment as it is used to capture these metaphysical notions. Hence, I claim that a different logical notion should be used than the notion of entailment that is used to understand proofs and theories.

My reason for this is quite simple. For any of these metaphysical purposes, one needs more metaphysically substantial connections between antecedent and consequent than when thinking about theories or proofs. For example, both Moore and Dunn tried to develop theories of what is necessarily contained in the concept of a thing. They both formalise the idea that the property P is essentially internal to the concept of a thing i if and only if the following formula is true:

$$\forall x(x = a \rightarrow Px),$$

where $\rightarrow$ is some form of entailment. I think most philosophers would now think that it is essentially internal to the notion of being a platypus that it is a monotreme

https://doi.org/10.1017/9781009375283.002 Published online by Cambridge University Press

(a mammal that lays eggs). In particular, it is essentially internal to Pauline the Platypus that she is a monotreme, and so

$$\forall x(x = pauline \rightarrow Mx).$$

It was, however, an empirical discovery that Platypuses were monotremes. A theory that predates this discovery should not be closed under this entailment. Hence, whatever sort of logic of entailment that formalises the notion of essential containment in this sense is different from the one that is a universal tool for the closure of theories.

The case with regard to truthmakers is rather similar. In formulating his version of the so-called truthmaker principle, John Bigelow said:

> Whenever something is true, there must be something whose existence entails that it is true. The 'making' in 'making true' is essentially logical entailment. [15, p 125]

Formalising this notion, what Bigelow said is that m is a truthmaker for p, then

$$m \text{ exists} \rightarrow p \text{ is true.}$$

Clearly, we should have a truthmaker version of Tarski's principle T, that is, for example,

The fact exists that water is wet → *water is wet* is true.

I use 'fact' and 'truthmaker' here as synonyms. Now, let's consider the following facts:

water is wet $\qquad H_2O$ is wet

On most current philosophical accounts, I think, these are the same fact. Hence, we should have

The fact that H_2O is wet exists → *water is wet* is true.

Once again, the sort of entailment needed here is too metaphysically loaded to provide an adequate treatment of the closure of theories. For that task, an entailment that can make much more fine-grained distinctions is required.

There are, therefore, at least two distinct notions of entailment: one that treats theory closure and proof plans and another that deals with metaphysical closures of various kinds.

1.6 Problem: Use and Mention

Some philosophers have claimed that treating entailment as a connective is a mistake. W. V. O. Quine was perhaps the most prominent of those who made this objection. He said (using 'implication' in the same way that I am using 'entailment'),

> [C. I.] Lewis, [H. B.] Smith, and others have undertaken systematic revision of '⊃' with a view to preserving just the properties appropriate to a satisfactory relation of implication; but what the resulting systems describe are actually modes of statement composition – revised conditionals of a non-truth functional sort – rather than implication relations between statements. [171, p 32]

https://doi.org/10.1017/9781009375283.002 Published online by Cambridge University Press

Quine had a good point. To say that a statement B follows from a statement A is to posit an entailment relation that holds between A and B. The expressions '$A \to B$' and '$A \Vdash B$' have very different logical forms. In '$A \to B$', the statements 'A' and 'B' are subformulas and there is no reference to them as sentences. In '$A \Vdash B$', there is reference to the sentences 'A' and 'B' themselves and the notion of entailment represented is a binary property of sentences (see also [169]).

For example, according to Quine, it was correct to say

'Every dog is mortal' entails 'all Spitz dogs are mortal',

but it was incorrect to say

Every dog is mortal entails that all Spitz dogs are mortal.

Thus, it is a grammatical mistake to try to construct a logic of entailment [171, p 28].

In an appendix to the first volume of their book, *Entailment*, Alan Anderson and Nuel Belnap reply to Quine by arguing that there is no grammatical error in treating 'entails' as a connective. They point out that a relation on sentences can be converted in ordinary English to a sentence operator rather easily. Their example of a monadic sentence predicate that can be converted to a sentence operator (i.e. to a connective) is quite clear.

'Tom is tall' is true.

In this sentence, 'is true' is a predicate, which holds of the sentence 'Tom is tall'. The string of words, 'Tom is tall is true', is not a sentence. For one, it has two main verbs. The expression 'is true', however, can be converted rather easily into a connective:

That Tom is tall is true

This is a well-formed sentence. In the expression, 'that _ is true', one can replace the blank with a sentence and obtain a sentence. Thus, it is an operator that takes sentences to sentences, that is to say, it is a connective [4, pp 479–480]. Even if 'entails' is a relation between sentences, in the same manner it can easily be converted into a connective:

That every dog is mortal entails that all Spitz dogs are mortal.

The expression 'that _ entails that _' is a binary connective.

Another approach to dealing with Quine's complaint is to claim that entailment is a relation between *propositions* instead of sentences. Thus, if we have a relation Ent such that '$Ent(A, B)$' means that the proposition that A is entailed by the proposition that B, then '$Ent(A, B)$' itself can express a proposition that can all be expressed by '$A \to B$'. As Quine said, 'entails' "would come to enjoy simultaneously the status of a binary predicate and the status of a binary sentence connective" [171, p 32]. Quine rejected this idea on the grounds that the entities needed (i.e. propositions) are too obscure and should not be postulated.

Alasdair Urquhart replied to this latter worry in his article, "Intensional Languages via Nominalization" [206]. In that paper, Urquhart shows how to construct a semantics for a modal logic in which '$\Box$' is taken to mean 'it is a tautology that'. The

https://doi.org/10.1017/9781009375283.002 Published online by Cambridge University Press

resulting logic is called TS for "Tautology System". In the syntax and semantics of TS, sentence names are *nominalised*. This means that, when they are arguments of the representations of expressions like 'it is necessary that' and 'entails', they are numerals that represent the Gödel numbers of the expressions that they nominalise. So, 'It is necessary that A' is represented as $\Box g(A)$, where $g(A)$ is the (numeral representation of the) Gödel number of A. Unless Quine wished to claim that numbers are obscure entities, he would have had to accept that Urquhart's construction allowed 'entails' and other intensional idioms to enter into our logical language.

I think, however, that Quine had another worry in mind in accusing intensional logicians of confusing use and mention. The logicians whom Quine was attacking, C. I. Lewis and Rudolf Carnap in particular, thought of their necessity operators as representing logical truth. For them (in more modern notation), '$\Box A$' meant 'A is logically true' or, to follow Anderson and Belnap, 'It is logically true that A'.

For Quine, logical truth was just generality.[4] Quine said that "a logical truth is a statement which is true and remains true under all reinterpretations of its components other than the logical particles" [168]. I think that, for Quine, logical truth was not primarily a property of sentences, but rather of schemes. A scheme is a formula that contains metavariables representing grammatical particles such as statements. Consider the scheme, $A \equiv B$. We cannot say that '$A \equiv B$' is logically true, logically false, or logically contingent, because it does not have any truth value at all unless 'A' and 'B' are given an interpretation. By themselves, the expressions A and B have no truth values. To deal with this problem, Quine introduced his quasi-quotation marks. The expression, $\ulcorner A \equiv A \urcorner$, for example, is true for all its substitution instances and so is logically necessary, and $\ulcorner A \equiv B \urcorner$ is logically contingent because some of its substitution instances are true and others are false [171, p 35].

In order to use Urquhart's device to reply to Quine's worry, we would have to introduce Gödel numbers for schemes. This would require a formalised metalanguage (with a finite primitive vocabulary). It would be possible to do this. Instead of having infinitely many "atomic" schematic letters, $A, B, C, \ldots$, as we usually do in informal metalanguage presentations, we could have a single basic letter A with superscripted primes – $A_{\prime}, A_{\prime\prime}, A_{\prime\prime\prime}$, etc. Then what we get, if the appropriate semantics is given to this language, is a logic with some implicit propositional quantification. The sentence $\Box g(\ulcorner A \urcorner)$ says that the scheme A is logically true, and hence that all instances of A are logically true. This seems rather complicated, but it seems doable.

Another, and I think better, way to avoid Quine's problem is to reject his view that schemes, rather than sentences, are the real logical truths. We can still maintain that a sentence is logically true only if all of its substitution instances (for its so-called non-logical particles) are true. Schemes are convenient means of expressing logical truths, but they are not the primary bearers of logical truth. In his "Syntactic Construction of Systems of Modal Logic", J. C. C. McKinsey treated modality in just this way. He evaluated formulas relative to a set of substitutions. A substitution associates with each propositional variable a formula of classical propositional calculus (i.e. one that

[4] Russell held this too (see Section 2.2).

does not contain any modal operators). The formula $\Diamond p$ is true because there is at least one formula that could be substituted for p that is true. And $\Box(p \vee \neg p)$ is true because no matter what formula is substituted for p, $p \vee \neg p$ comes out true [46, 140]. We could interpret entailment in this way too. We could say that $A \rightarrow B$ is true if and only if for every substitution that makes A true also makes B true. The result is a normal classical modal logic. I argue in Chapter 3 that this sort of logic is inadequate to represent logical entailment, but not because it runs afoul of any strictures governing use and mention.

I do not think, however, that all this formal machinery is really necessary. Quine is right that '$A \Vdash B$' mentions the formulas A and B in a way that '$A \rightarrow B$' does not. But, given the right semantics we can see that there is a correlation between true entailments and valid deductions, and this semantics should explain why there are these correlations. In this way, entailments do express the validity of sequents and the use-mention problem is a pseudo-problem.

This representation of elements of the metalanguage in the object language raises other interesting issues. We can make explicit that we are talking about sentences if we add names of sentences to the language and represent entailment as a predicate rather than a connective. '$E(x, y)$' could mean 'x entails y', where x and y range over sentences. As I discuss in Chapters 5 and 10, the explicit addition of names for sentences together with certain widely held logical theses, can lead to difficulties such as Jc Beall and Julian Murzi's "validity Curry paradox" [12].

1.7 Logicality

As I say in Section 1.6, Quine thought of logical necessity as generality. Although I do not think that logical necessity is merely generality nor do I think that entailment is reducible to universally true implication, I do think that generality has something to do with logical entailment. A logic of entailment, in my view, has to be *formal*, but what exactly this means needs to be spelled out in some detail.

In his PhD dissertation, John MacFarlane set out three different senses of the claim that logic is formal. This taxonomy has become quite influential and so I think it will be helpful to use it to situate my own view. These were MacFarlane's three senses of 'formal' [121, ch 3]:

1. Logic is formal in the sense that its norms are "constitutive of thought as such";
2. Logic is formal in the sense that it is "indifferent to the particular identities of objects";
3. Logic is formal in the sense that it "abstracts entirely from the semantic content of thought".

The first sense of 'formal' is to be found in Kant's writings. Kant said that, for example, a thought that violated the law of non-contradiction was not actually a thought. Thinking would cease to be thinking if it violated any laws of logic. The third sense could be found in Kant and also in the Wittgenstein and the logical positivists.

https://doi.org/10.1017/9781009375283.002 Published online by Cambridge University Press

Wittgenstein and the Positivists held that the laws of logic had no semantic content. This view is incorporated into Carnap and Bar Hillel's theory of information too. On that theory, the amount of information carried by a proposition was proportional to how implausible it is. A contradiction carried the maximum amount of information and a tautology carried no information at all.

The second sense of 'formal' is the one that I attribute to the logic of entailment. The idea is very old. It was held by several mediaeval philosophers, in particular by Peter Abelard and Jean Buridan. Both Abelard and Buridan were trying to elucidate Aristotle's distinction between perfect and imperfect deductions [6, 24b23]. Aristotle called a deduction perfect if nothing had to be added to it to make evident that the conclusion followed from the premises. Abelard and Buridan give similar examples of imperfect deductions. One of Abelard's examples was "If every man is an animal, every man is alive" [121, p 281] and one of Buridan's was "A human is running, so an animal is running" [27, ch 14].

Abelard's gloss on the imperfection of these sorts of inferences was that they relied on the nature of the subjects and predicates involved rather than just on the logical form of the expressions. Buridan said that an imperfect deduction was valid in a sense because it was impossible for its premise to be true and conclusion false, but it is flawed in that not every deduction of the same form is valid [27]. If we substitute 'rock' for 'human' in Buridan's example, the argument is no longer valid even in the weaker sense.

I think that the logic of entailment should be topic neutral in this way for pragmatic reasons. I do not wish to treat conditionals such as 'if x is water, then it is H_2O' as logical entailments. Taking them to be entailments can interfere with one of the jobs I think is important for the logic. A logic of entailment is, among other things, a theory of how theories should be closed. Forcing all theories to be closed under these sorts of analytic conditionals may interfere with the content of theories. We do not want to close a theory about phlogiston in such a way that it talks about oxygen. For this reason, I think that the logic of entailment is formal in MacFarlane's second sense.

As John Etchemendy [59] pointed out, it was the hope of some logicians that by designating a set of "logical constants" (the usual connectives and quantifiers), they could give a reductive analysis of logical truth and entailment in terms of actual general truth. For example, 'every cat is either a cat or a dog' is logically true because its generalisation, $\forall F \forall G \forall x(Fx \supset (Fx \vee Gx))$ is true. I suggest this was true of Quine (see Section 1.6) and Russell (Section 2.2), and Etchemendy pointed out that it was in Tarski as well. Etchemendy argued, successfully I think, that demarcating logical from non-logical vocabulary alone could not provide a successful reductive account of logical as generality, nor could it provide the basis for a logic of entailment.

Etchemendy suggested, moreover, that the choice of certain pieces of vocabulary as logical was arbitrary. I think that Etchemendy was right that, if we rely on our intuitions about logical necessity and derivability alone, then it seems there is nothing special about a *formal* necessity such as '$\forall F \forall G \forall x((Fx \wedge Gx) \supset Gx)$' as against an informal necessity such as 'every red thing is coloured'. I think the problem here, however, is that Etchemendy uses as his only benchmark of truth the compliance of a theory with a set of intuitions about logical truth and logical consequence. I use as evidence for the correctness of my view its ability to do certain tasks that a logic of entailment is supposed to do.

These tasks are to act as a general theory of theory closure and to provide a logical tool to analyse proof plans. To do the latter, I argue, we need an entailment connective. To do the former, we need to represent not just what each sentence in a theory entails, but also what a collection of sentences jointly entail. As we shall see, representing the closure of theories requires that we have a conjunction that has certain properties. Thus, the logic has to have an entailment and a conjunction, and both of these need to have certain properties (discussed in the proceeding sections). I put forward a logic with just these two connectives and I call it the "core logic of entailment" (see Chapter 8).

I hold that all theories should be thought of as closed under the logic. In order to do so, the logic has to be indifferent to actual and even some metaphysically necessary features of things. So, given the need for substantive principles governing entailment and conjunction (at least) in order to provide an adequate logic of theory closure and the need to be indifferent to the metaphysics of the subject matter, there is a distinction that we can make between the logical and non-logical vocabulary in statements.

1.8 Entailment and Necessity

Since the dawn of the study of logic in the Western tradition at least,[5] logicians have thought of the notion of logical consequence as incorporating some notion of necessity. In the *Prior Analytics* Aristotle wrote that

> A deduction is a discourse in which certain things being stated, something other than that which is stated follows of necessity from their being so. [6, 24b18]

Apart from those logicians who question the very coherence of the notion of necessity, such as Russell and Quine, it has been accepted universally that logical consequence incorporates some notion of logical necessity. Even in introductory lectures on logic we appeal to some form of logical necessity. We typically tell students that a valid argument is one in which it is impossible for the premises to be true and the conclusion false. I follow the mainstream in the theory of entailment by claiming that true entailment sentences reflect the necessity of the corresponding valid deductions, that the entailment connective embodies this sort of necessity. What this necessity consists in, however, is quite a difficult issue.

Many philosophers equate logical necessity with metaphysical necessity. Perhaps, the most vocal modern proponents of this view were Ludwig Wittgenstein (in the *Tractatus*), David Lewis [115], and Frank Jackson [88]. There are, however, other views of the necessity involved in logical consequence. Michael Dummett viewed

[5] There is also a debate about the existence and nature of logical necessity in Classical Indian philosophy but I am certainly not enough of an expert to contribute in any way to that debate. So I do not discuss it here.

https://doi.org/10.1017/9781009375283.002 Published online by Cambridge University Press

logical consequence in terms of the transmission of justified assertion. We can understand the sort of necessity that he postulated as a form of epistemic necessity. Another interpretation of logical necessity has been suggested by Greg Restall. On Restall's view, a classical sequent,

$$A_1, \ldots, A_n \Vdash B_1, \ldots, B_m,$$

is to be understood as saying that one should not simultaneously assert all of $A_1, \ldots, A_n$ and deny all of $B_1, \ldots, B_m$ [179]. In the sort of inferences that I am concerned with in this book, where there is only a single conclusion, a sequent $A_1, \ldots, A_n \Vdash B$ is said to be valid if and only if one should not simultaneously assert all of $A_1, \ldots, A_n$ and deny B. In this view, norms that are represented by valid sequents express a notion of deontic necessity. I discuss this view in more depth in Section 3.8.

In Chapter 3, I discuss attempts to base a theory of entailment on a theory of logical necessity and possibility. In particular, I discuss attempts to give a possible worlds semantics for logical necessity and treat entailment as truth preservation on all logically possible worlds. I argue there that such approaches do not give an adequate treatment of nested entailments. Even though I find this sort of worlds treatment of entailment wanting, I do think that any adequate theory of entailment (like any adequate theory of logical consequence) does need to capture some notion of logical necessity. In Chapter 8, I discuss the relationship between the logic of entailment and logical necessity and argue that the rather weak sense of necessity captured in the logic is adequate.

1.9 The Principle of Inclusion

I hold that there are four core principles of entailment. They are the principles of inclusion, monotonicity, transitivity, and compactness. Each of these is controversial, so I will spend a few words formulating and defending them. I start with the principle of inclusion. (This principle is sometimes called the "principle of reflexivity", but I use "reflexivity" for another property, see Chapter 5.)

The principle of inclusion says that any sentence in a set of sentences is a logical consequence of that set, follows from that set. In formal notation, the principle of inclusion is

$$\frac{A \in \Gamma}{\Gamma \Vdash A}.$$

The reason for accepting inclusion is quite obvious. In a proof about a theory, one often makes use of the axioms or other known theorems of that theory. In so doing, one is in effect appealing to the principle of inclusion.

There have been, however, logicians who have rejected inclusion. Robert Meyer and Errol Martin said:

https://doi.org/10.1017/9781009375283.002 Published online by Cambridge University Press

> Which arguments are valid? This has been the central question of logic. "Reasoning is an argument in which, certain things being laid down, something *other than these* necessarily comes about through them," said Aristotle (*Topics* 100a 25–27). The emphasis is ours. "He who repeats himself does not reason," as Strawson correctly notes. The fallacy of concluding what one has assumed is almost universally condemned. Some of the rubrics under which it is condemned are the following: *circular reasoning, begging the question, petitio principii.* [145]

If we accept inclusion, we also have to accept that the proof plan, $A \to A$, is valid. Meyer and Martin said that we should not accept a proof plan of this form. This form is the form of what has traditionally been thought of as a logical fallacy – begging the question.

At this point, it may seem that I have stumbled upon a place in which the logic of proof plans diverges sharply from the logic of theory closure. It is clear that in thinking about theory closure the principle of inclusion is required. Given some statements from a theory, such as its axioms (if it has any), one can infer various things about what that theory contains, including those statements themselves. But a proof plan of the form $A \to A$ might seem useless. In fact, such proof plans may seem harmful. We think it is important that when proving something the conclusion should not appear in the premises. One who sneaks his conclusion into his premises would be called a charlatan of some sort, or at best declared to be sloppy. I think, however, that this divergence between the logics of proof plans and theory closure is merely apparent rather than real.

Consider the following example. Suppose that there is a logical conflict between A and B, that is, it is true that $\neg(A \wedge B)$. Suppose also that we know that it is a theorem that A entails B. Then, we know that A cannot be true, that is $\neg A$ obtains. One way of representing the inference this is:

$$\frac{A \Vdash B}{\neg(A \wedge B) \Vdash \neg A}.$$

This seems a good inference. What is going on here, I maintain, is that once we know that B follows from A, we also know that $A \wedge B$ follows from A, that is

$$\frac{\dfrac{A \Vdash B}{A \Vdash A \wedge B}}{\neg(A \wedge B) \Vdash \neg A}.$$

How do we know that the argument above is valid? One good reason is because the following inference is obviously valid:

$$\frac{\dfrac{A \Vdash A \qquad A \Vdash B}{A \Vdash A \wedge B}}{\neg(A \wedge B) \Vdash \neg A}.$$

We tacitly assume the validity of $A \Vdash A$ in this and many other inferences. We do not often use inclusion overtly, but we do sometimes use it tacitly.[6]

[6] A referee correctly pointed out that there may be many ways to derive that $A \Vdash A \wedge B$. But, note that if we accept that any of these derivations is valid, then we can infer that $A \Vdash A$ from $A \Vdash A \wedge B$ and conjunction elimination. Thus, if every formula implies some formula or other, then inclusion is valid. The point here is that it really is very difficult to avoid making inclusion valid.

1.10 Monotonicity

The principle of monotonicity says that if a statement follows from a set of premises, then it still follows if we add further premises. Formally stated, this is what the principle of monotonicity says:

$$\frac{\Gamma \Vdash A \qquad \Gamma \subseteq \Delta}{\Delta \Vdash A}\,.$$

Recall that in '$\Gamma \Vdash A$', we think of the premises (the members of Γ) as being conjoined to one another. As I say in Chapter 4, in relevant logic there is a sense in which the premises of an inference are not treated as being conjoined to one another and on this notion of inference, monotonicity fails. For the present discussion, however, I put aside this relevant notion of consequence ($\vdash$) and discuss only the relation, $\Vdash$.

There is an important challenge to monotonicity that takes the most general form of reasoning to be a type of counter-factual reasoning. Counter-factuals are not monotonic. It does not follow from 'If Nova had been good, then Sue would have been happy' that 'If Nova had been good and there were a major earthquake in Wellington, then Sue would have been happy'. Daniel Nolan puts forward a view of this sort [154, 155]. He suggests that reasoning about, for example, inconsistent situations is like reasoning counter-factually. He says:

> imagine I reasoned by saying "suppose naive set theory were true. Then the Russell set is a member of itself and not a member of itself. So at least one set is a member of itself and not a member of itself". One way to understand what has happened is that the effect of the reasoning is much the same as before: I have not categorically supported the conclusion that at least one set is a member of itself and not a member of itself: instead, the force of my conclusion is captured by "if naive set theory were true, at least one set is a member of itself and not a member of itself". [155, p 422]

On Nolan's view, there is no general logic of entailment. If one considers the hypothesis that some alternative logic holds, then one cannot use any of her usual logical principles that are not contained within this alternative system. This view of deduction and the correlated view of entailment do not allow for a single logic of entailment (except for the logic with no generally valid principles).

Thinking about deduction in this non-monotonic fashion, however, clashes with our ordinary notion of proof. Let's say that one proves something from some (but not all) of the axioms of a theory. If this is an actual *proof*, then the mention of further axioms cannot show that the conclusion is false. But, if the relation of provability is non-monotonic, then we can undermine conclusions in this manner. The notion of proof that I am discussing, then, requires that entailment be monotonic.

I return to the subject of non-standard logical theories at the end of Chapter 8. There, I suggest that we can retrench our logic in terms of the negation and disjunction principles that we accept when faced with certain non-standard theories. In *Relevant*

Logic, I suggest that when faced with the need to reason in a very non-standard way, we should think of this in a metalinguistic manner. That is, our reasoning in such situations is in effect in the scope of an "according to the logic such-and-such ..." operator. While reasoning in the scope of such an operator, only the inferential moves licensed by the logic mentioned are legitimate. But such reasoning is, in a sense, playing a formal game – one manipulates the symbols of the language in the way that is set down by the proof theory or by the model theory of the logic. In different situations, I think that it is appropriate to choose one and not the other of these solutions, but I cannot see that there are situations in which neither is adequate. Thus, I accept a monotonic notion of proof.

1.11 Transitivity

The third constraint that I place on a consequence relation is *transitivity*:[7]

$$\frac{\Gamma \Vdash A \qquad A \Vdash B}{\Gamma \Vdash B}.$$

In Chapter 6, I give an argument for transitivity in this form and in Chapters 7 and 8 I give arguments for stronger forms of transitivity. These arguments have to do with closure operators used to construct theories. This is a technical topic and is better left until after the required formal background is discussed. Here, instead, I briefly present an intuitive and informal motivation in favour of transitivity broadly construed.

Our notion of proof seems to include a notion of steps towards a proof. In reasoning about empirical theories, we might have an entailment between a theoretical postulate and a possible observation. In mathematical proof plans, steps in reasoning are axioms or lemmas. When we derive a consequence from a set of axioms, say, we add it to the axioms and allow it to be used in further proofs. A theory is what we construct, or rather what could be constructed in ideal circumstances, out of all the consequences drawn in this way. A consequence that is unwanted is a purported counter-example to the theory. The fact that it is done in one step or in two hundred steps makes no difference to its status as a counter-example.

The concept of a step in a proof seems integral to the structure of proof plans. We think of these steps to be taken one after another in a logical progression towards the desired conclusion. Sometimes the way in which lemmas are put together to derive the conclusion is not sequential or the proof has a more complicated structure than the transitivity rule given above, but that does not undermine the claim that this simple rule is required by our ordinary notion of proof.

[7] David Ripley claims that the property that I set out is not really a transitivity property, but a "linking property" that he calls "$\mathrm{KS}_{\mathrm{SF}}$" [181]. It does not matter for my purposes whether this is referred to as transitivity rule or a linking property.

1.12 Compactness

One of the most important properties a consequence relation must have in the context of the current project is *compactness*. A consequence relation, $\Vdash$, is compact if and only if for every set of formulas Γ and every formula A, if

$$\Gamma \Vdash A,$$

then there is some finite subset of Γ, Γ', such that

$$\Gamma' \Vdash A.$$

The reason that this is so important is that the connectives of the language that I use are finitary. I represent the derivability of A from the set Γ by the entailment of A by some conjunction of formulas of Γ. If there is no finite set of formulas in Γ from which A can be derived, then I cannot represent the derivation of A from Γ in this way. I discuss the possibility of constructing an infinitary logic of entailment in Chapter 6, but the central point is that human inference is in an essential sense, finitary. I do think we can *appeal to* infinitary rules, but we cannot use them. We might discuss a version of the theory of arithmetic that is closed under the infinitary omega rule, for example. We cannot use the omega rule, but we can appeal to it to say, for example, that Gödel's theorem is not provable of this theory. I make clear this distinction between using a rule and appealing to it in Chapters 6 and 11.[8]

Another way of dealing with the issue of the failure of compactness is to add names for sets of formulas to the language. Let γ be a name for a set of formulas, Γ. Then, we could represent $\Gamma \Vdash A$ by $K(\gamma) \to A$, which means that the conjunction of formulas in Γ entails A. Although this avoids the move to an infinitary language, it still has problems. If there are $\aleph_0$ many formulas, then (assuming the continuum hypothesis) there are $\aleph_1$ many sets of formulas, and hence we would need at least $\aleph_1$ many names in the language. If there are $\aleph_1$ many names, then there are at least $\aleph_1$ many formulas that can be constructed from them and so there would then be $\aleph_2$ many sets of formulas. Then, we would need $\aleph_2$ many names. And so on, up the hierarchy of the transfinite cardinal numbers. Now, suppose that $\Gamma \Vdash A$ if and only if $K(\gamma) \to A$ is a theorem of the logic, for all sets of formulas Γ. Then, there is something interestingly finitary about the logic: for every valid sequent, there is a finite sequent (and a valid formula) that represents it. This is an interesting approach, but difficult to formulate and it is essentially unaxiomatisable. And so I set it aside.

For these reasons, I place the constraint consequence relations that are to be considered as the basis for a theory of entailment that they be compact.

[8] Sometimes this property is called "finitariness" rather than "compactness", where the latter term is reserved for the property that a model or class of models have according to which a set of formulas is satisfiable if and only if every finite subset is satisfiable. Classically, finitariness and compactness are equivalent, but they are not equivalent in all non-classical frameworks [86, ch 1].

1.13 Method

An entailment $A \rightarrow B$ means that B is deducible from A. So, $A \rightarrow B$ is true if and only if B is deducible from A. Any adequate theory of entailment has to make plausible that this biconditional obtains. It may seem very straightforward to do this, but in fact it can be rather tricky. Consider, for example, a theory that claims that the deducibility relation of the modal logic S5 is the one true deducibility relation. On this theory, $A \rightarrow B$ is true if and only if B is deducible from A is S5. What we need here is a model in which all and only the S5 valid inferences are reflected in true entailments. We cannot use any of the usual models for S5. Consider one such model and pick a random world w to be the actual world. Let p be a formula that is true in w. Then, at w, the entailment

$$\Box\neg p \rightarrow B$$

holds for every formula B. The corresponding deduction,

$$\Box\neg p \Vdash B,$$

however, is not valid in S5 for all formulas B. So this is not a model in which every true entailment expresses an S5-valid deduction. As I explain in Chapter 3, those who treat entailment in terms of a possible worlds semantics have tended to accept extensions of S5 as the logic of entailment.

In Chapters 7–10 and the Appendix, I argue that there is a model (the "intended model") that makes all and only true entailments true. And this is a model for the logic E.

The development of a model theory for the logic also has another goal. In my theory, the model theory is a semantics for the proof theory. The model theory explains the proof theory. In explaining the proof theory, the semantics also justifies the proof theory. The semantics demonstrates the rationale for the various rules of proof. But the fact that the proof theory is elegant and to a large extent intuitive also gives some justification to the model theory. It is justified by being a semantics for that proof theory.

Although intuitiveness is an important criterion for the choice of a logic of entailment, it is not the only criterion. Other criteria include:

1. The entailment connective of the logic must express a deducibility relation that has plausibility independent of being expressed by the connective;
2. The treatment of nested entailments must be reasonable;
3. The logic must prove as theorems the formal representations of paradigm cases of good proof plans.

These three criteria have to do with the logic being able to do the job of a logic of entailment. As such, our choice of a logic of entailment cannot fail to have any of these three properties. If a logic does fail to do any of these jobs, then it is not really a logic of entailment. A logic of entailment has jobs to do and it had better do them.

One might find my adherence to criterion three surprising, because I am an advocate of relevant logic. No relevant logic contains all instances of the following scheme:

$$((A \vee B) \wedge \neg A) \rightarrow B.$$

But, surely (an objector might say), the following is a paradigm of a good form of a proof plan: if I can prove $A \vee B$ and $\neg A$, then I can prove B. I argue in Chapter 11 that this proof form, that is disjunctive syllogism, is to be understood not as a universal proof form but only as a rule used in the construction of some theories. In making inferences about those theories, we can appeal to disjunctive syllogism, but we cannot do so in making inferences about theories in general. I leave the argument for that view for later, but here I would like to make a point about method. Denying that a particular rule – even a widely accepted rule – is a good deductive principle is allowed in the search for a logic of entailment as long as one can explain why that rule seems falsely to be a good rule. In the case of disjunctive syllogism, the fact that the creation of consistent theories is a widely accepted goal of theory creation in the formal and empirical sciences makes the appeal to disjunctive syllogism so common that it seems to some to be a generally applicable rule of inference.

Satisfying the criteria listed above does not uniquely determine a logic of entailment. We need to appeal also to the theoretical virtues such as elegance, strength, and, of course, intuitiveness, in deciding what logic to accept. This may leave some room for a form of logical pluralism about entailment. I have chosen, however, to avoid discussing the topic of pluralism in this book because the literature on pluralism is now very extensive and engaging it would mean writing a very different book than the one I have in mind.

1.14 Plan of the Book

The book is divided into three parts. Part I gives a history of logics of entailments. It is not meant to be a complete history of the subject. Rather, it looks at some of the more important and influential approaches to entailment and in particular those that influenced me in my adoption and interpretation of Anderson and Belnap's logic E. Anderson and Belnap's logic is itself the subject of one of the chapters in Part I. This inclusion, of course, introduces the logic, but it also argues that their logic was, despite their proof theory, in need of further interpretation. In Part II of the book, I introduce Tarski's consequence operators and then employ them to construct a model for a simple logic that I call Generalised Entailment Logic (GTL). The idea is to use the notion of theory closure, which is what is produced by a consequence operator, to act as the basis for the semantics of entailment. Part III begins with an argument that GTL is too weak to be treated as the logic of entailment. In Chapter 8, a semantics for the implication and conjunction fragment of E is presented, and in Chapters 9 and 10, this semantics is extended to treat all of propositional and first-order E. In the final chapter, I look at the integration of E and its semantics into a more general view of deductive inference.