PART I

THE PROBLEM
OF
CONDITIONALS
I.1. METHODOLOGICAL REMARKS

To those who believe that there is such a thing as the logic of conditionals this dissertation may appear to be yet another attempt to unravel its secrets - which it is not. As a logician, you can do no more than devise a logic for conditionals and try to persuade your readers to adopt it. You may succeed in doing so if you are able to demonstrate that the one you propose is a better logic for conditionals than the ones proposed so far. The phrase 'better for conditionals' should, however, not be misunderstood. In particular, it should not be interpreted as meaning 'more like the real one'. The best logic for conditionals one might propose is not that which they actually possess. It cannot be, not because this actual logic would be not good enough, but simply because there is no such thing. Whether a given logic is better than some alternative has little to do with its better fitting the facts; it is more a question of efficacy.

This is one of the theses defended in the following introductory pages. It is put forward when the question is discussed as to how one should choose between rival logical
theories. That is a very natural question to ask in an introduction to the problem of conditionals. If only because so many theories have been put forward, all purporting to solve the problem, that putting yet another one on the market might seem to only add to the difficulties.

I.1.1. The case of the marbles

Here is an example which will return regularly in much of the following. There are three marbles: one red, one blue, and one yellow. They are known to be distributed among two matchboxes, called 1 and 2. The only other thing which you are told is that there is at least one marble in each box.

The various possibilities which this leaves open can be summarized as follows:

<table>
<thead>
<tr>
<th></th>
<th>box 1</th>
<th>box 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>blue</td>
<td>yellow, red</td>
</tr>
<tr>
<td>II</td>
<td>yellow</td>
<td>blue, red</td>
</tr>
<tr>
<td>III</td>
<td>red</td>
<td>blue, yellow</td>
</tr>
<tr>
<td>IV</td>
<td>yellow, red</td>
<td>blue</td>
</tr>
<tr>
<td>V</td>
<td>red, blue</td>
<td>yellow</td>
</tr>
<tr>
<td>VI</td>
<td>blue, yellow</td>
<td>red</td>
</tr>
</tbody>
</table>

Bearing these possibilities in mind, you will have to agree that

(1) The yellow marble is in box 1, if both of the others are to be found in box 2.

Now suppose someone no better informed than yourself were to claim that
(2) The yellow marble is in box 1 if the blue marble is in box 2

You will disagree. You may even go as far as to assert

(3) It is not so that if the blue marble is in box 2, the yellow one is in box 1

After all, the yellow marble could just as well be in box 2 together with the blue one, as long as the red one might yet be in box 1. Things are different, however, if it is excluded that the red marble is in box 1. In that case, (2) holds. That is,

(4) If the red marble is in box 2, then if the blue marble is in box 2 as well, the yellow marble is in box 1

For those who do accept the last two statements there is a surprise in store. Using standard logical notation 1), we see that sentences (3) and (4) are respectively of the form

(3') \( \sim [\text{blue in 2 } \rightarrow \text{yellow in 1}] \)

(4') \( \text{red in 2 } \rightarrow [\text{blue in 2 } \rightarrow \text{yellow in 1}] \)

So we can apply the principle of Modus Tollens to (3') and (4') and conclude

(5') \( \sim \text{red in 2} \)

That is,

(5) It is not the case that the red marble is in box 2

But this conclusion will be a lot less acceptable than the premises (3) and (4) might have seemed. 2)
Is Modus Tollens playing up here, or should (3) or (4) be rejected after all? Any who have been confused by the above will feel obliged to choose sides.

Some may choose to stay with their initial intuitions about (3) and (4), regarding the example as evidence that Modus Tollens sometimes fails. But what kind of evidence is this? Why should intuitions be treated with this kind of respect, especially where others do not share them?

Choosing instead to do something about (3) and (4), and thus save the principle of Modus Tollens, would seem a lot safer: all existing theories of conditionals will support this. Even so, one would not have an easy time working things out. For although everyone is agreed that something must be done, there is no consensus to be found in the literature as to precisely what it should be. Some theories will advise you to reject (3), other ones to reject (4). Often, however, the response is more sophisticated, amounting to a denial that we are dealing with a proper instantiation of Modus Tollens here. Thus it can for example be argued that (3) is incorrectly formalized as (3'), as a negation of an implication; there is a hidden operator, and (3) really means something like 'it is not necessarily the case that if the blue marble is in box 2, the yellow one is in box 1'. Or it is argued that (4) is incorrectly translated into (4') as an embedded conditional sentence, that its meaning lies closer to something like 'if both the red and the blue marbles are in box 2, then the yellow one is in box 1'.

We will be returning to these matters at length in chapter 2, so we save the discussion up for there. Suffice it here to remark that it is not obvious how this partiality for Modus Tollens is to be justified. As above, by an appeal to intuition?
I.1.2. Logic as a descriptive science

The marbles puzzle has not been presented under the illusion that it would refute any theory of conditionals developed so far. Actually, it was presented under quite a different illusion, namely that it might help to put the aims and methods of one at present quite popular approach to the logical analysis of natural language in an unfavourable light.

Typical of this approach, in what follows to be known as Methodological Descriptivism, are
(I) a drive towards descriptive theories; that is attempts to distinguish systematically between the valid and invalid arguments within some class of arguments taken from a given natural language;
(II) the idea that such a theory can and should be tested by comparing what it has to say about the validity of the arguments it covers with the intuitive judgments of those who use the language concerned.

Key word in (I) is 'descriptive'. It is supposed that a logical theory should be descriptive - and that it can be so, the idea being that the logical researcher is faced with a class of arguments, some of which are, in fact, valid and some of which are, in fact, invalid. The object of the investigations is to discover where the division lies and, if possible, to find out why it lies there and not elsewhere.

'Descriptive' appears in the philosophy of logic in senses other than this, as a qualification not of logical theories but of the logical laws they sanctify. For instance, it sometimes crops up in discussions about the relation between logic and reality. Do the laws of logic tell us anything about the world? 'Yes, they do' is the descriptivist answer: the laws of logic add up to a compendium of the broadest traits of reality, and, in the last resort, owe their validity to their correctly describing these. One also frequently encounters the notion
in discussions about the relation between logic and thought. There, a descriptivist is someone who tries to ground the validity of the laws of logic in the actual mental process of reasoning: 'logic is the physics of thought, or it is nothing'.

It is quite possible to be a descriptivist in the methodological sense while having nothing to do with either of these outmoded convictions about logical validity. Thus, the strictest of conventionalists can hold that the validity of the laws of logic is just a result of the linguistic conventions which govern the use of the logical constants, that other conventions about the use these words would yield other logical laws, and from this we may conclude that the laws of logic are not descriptive, neither of reality (not even of its basic structure\(^5\)), nor of mental process (not even of those occurring in the soundest mind)\(^6\). Still, a conventionalist must be rated as a methodological descriptivist if we insist that logicians, in their analysis of a given language, should restrict themselves to describing the conventions which happen to bind the speakers of that language, and that they should refrain from anything like reforming those conventions.

So much for (I). (II) must be seen as a first attempt to answer the following question: assuming that it is reasonable to demand that a logical analysis should result in a descriptive theory, then how are we to find out whether any such purported description really conforms to the facts? Admittedly, (II) calls for some elaboration. What are intuitive judgments, and why do they matter when a logical theory is put to the test, while non-intuitive judgments do not? Two different answers can be distinguished in the literature, a traditional and a modern one.
I.1.2.1. The rationalist tradition

To appreciate the traditional answer we must let ourselves be carried back to the time when nobody ever seriously reckoned with the possibility of there being various alternative logics. At that time, no logician engaged himself in so many words with 'the logic of natural language'. One was, as it were, in search of the one and only logic (or believed that it had already been found in Aristotle's Syllogistics or Frege's Predicate Calculus), and that one logic was as a matter of course taken to be the logic of the natural languages.

Furthermore, it is important to realize that until about sixty years ago, semantics was of no more than marginal importance to the discipline of logic. Until then the core of a logical theory was given by a system of deduction, comprising axioms and rules of inference. These were couched in a more or less artificial language, of which often not even the syntax, let alone the semantics, was explicitly stated.

Now, a minimal requirement for a system of deduction - for any system of deduction, if at least it is presented as a characterization of some logic - is that it be reliable: every sentence that can be deduced from a given set of sentences must really follow from that set.

Traditionally, a system of deduction - every system of deduction ever presented as a characterization of the one and only logic - was held to be as reliable as its various axioms and rules of inference, and their reliability was thought to be self-evident. Take, for example, the law of non-contradiction: it is not the case that both $\phi$ and not $\phi$; wouldn't it be perverse to deny its validity? Or take the principle of Modus Ponens: from $\phi$ and $i\phi$ then $\psi$ it follows that $\psi$; isn't it obvious that this is always the case, whatever sentences $\phi$ and $\psi$ may be? Such principles simply cannot possess any better credentials than their self evidence; they cannot be rationally justified, for they are
themselves the principles which any rational justification must presuppose.

Self evident axioms and rules of inference, if indeed they are self evident, will of course make any further analysis redundant. But how do we grasp the validity of these principles of reason if reason itself cannot help us out?

It is there that intuition comes in: if our normal intellectual faculties fail, something else must enable us to acquire knowledge of the principles of reason. (Actually, it must enable us to do so a priori, i.e. long before reason will ever feel the need to exploit them.) This idea of a special intellectual faculty - intuition - may sound rather ad hoc. But its significance for the history of philosophy can hardly be overestimated. Indeed, it is not too much to say that from Plato onwards, philosophers, in particular those standing in the rationalist tradition, have constantly been trying to make it into something more than an ad hoc solution.

For our purposes, it is not necessary to treat these matters in further detail. It suffices to note that the development of various non-equivalent systems of deduction has made it increasingly impopular to establish the reliability of a logical theory in the manner described above. Indeed, even if one sticks to the absolutist view that there is only one logic, one must at least admit that any traditional estimation of the time and the trouble necessary to find it would be overoptimistic. Self evidence has turned out to be an unworkable criterion.

In discussions of natural language various authors may, nevertheless, still regularly be heard advocating one or the other argument form as being intuitively valid. Even non-absolutists tend to do so as soon as the discussion is restricted to the logic of a given natural language. This, however, is by itself not enough to convict them of the views here described. For, although this usage of 'intuitively' is certainly rooted in the traditional usage,
the word has in the meantime lost so much of its original impact that it might just as well be scrapped. Authors using 'intuitively' in such a context often mean no more than that the argument form in question seems reasonable enough to them, thereby not excluding the possibility that it might turn out to be invalid after all. In other words, in such a context 'intuitively' marks the introduction of an hypothesis rather than of an established truth.

Setting aside the descriptivist connotations, there is of course no fault to find with this usage of 'intuitive' just as long as one is aware that it would be begging the question to defend such an hypothesis by recourse to its intuitive validity if one is confronted with, for example, an argued counterexample. That would only lead to dogmatism. Or, whenever as in the case of conditionals, so many divergent views are held by so many authors, to an impasse.

I.1.2.2. The empirical approach

On to the second, more fashionable view of intuitive judgments. In practice it is not difficult to distinguish these from the first and traditional sort.

As we saw in the previous section, intuition was traditionally summoned in order to establish general logical principles. Consequently, traditional intuitive judgments - or at least those found worth recording - always say that some argument form, e.g. the Principle of Modus Tollens, is logically valid. Modern intuitive judgments, on the other hand, mainly serve the purpose of falsifying putative logical principles. They are judgments of concrete arguments (e.g. the particular instance of Modus Tollens given in section 1.1) which typically turn out to be intuitively 'absurd'.

Furthermore, 'intuitive' in the traditional sense goes with 'intuition', in the singular, sometimes preceded by 'our'. In the case of 'intuitive' in the modern sense, on the other hand, the plural 'intuitions' is employed, and mostly it is
not our intuitions which are at issue but those of the native speakers of the language concerned. If an author speaks of 'our intuitions', a precautionary 'pre-theoretical' or 'untutored' will usually be inserted.

Indeed, 'intuitive' in the modern sense is directly opposed to 'theoretical'. Any judgment arrived at by straightforwardly applying some logical theory to a given argument is deemed non-intuitive. Of course, any such judgment is of no use whatsoever if it is the reliability of the theory applied or of any of its rivals which is at stake. Only the judgments of the theoretically unbiased can then be allowed to count - this in order to preserve the impartiality, one might almost say the objectivity, of the data against which the predictions of the theory concerned are to be tested.

Hence, intuitive judgments in the modern sense are judgments about the validity of concrete arguments made by theoretically unprejudiced speakers of the language concerned.

There must be more to intuitive judgments than this. How reliable are they? After all, it would seem that even the most impartial arbiter may be mistaken in her judgment. Can we not hope for more trustworthy data?

Commonly 9) it is suggested that all competent speakers of a given language must, in virtue of their competence, be implicitly acquainted with its logical characteristics, and that it is this subconscious knowledge which surfaces in an intuitive judgment. Now clearly, any judgment betraying knowledge, even subconscious knowledge, must be correct. So, if this is what intuitive judgments do, then they are all true. Unfortunately, however, one cannot tell by just the form of a judgment whether it is a case of bona fide knowledge or merely one of belief. It is even impossible to decide whether ones own judgments were implicitly known to be true before they were explicitly believed to be so. Take the case of the marbles for example: do you know your own response to be the correct one, or do you merely believe this?
So this suggestion does not really help much. The gap between impartiality on the one hand and incorrigibility on the other appears unbridgeable, at least in practice. Or are we perhaps supposed to consult only speakers so competent that they are never mistaken? Then how should they be selected?

Some general criteria are of course readily available: very young children cannot be expected to have had the opportunity to develop their language skills, while others may for some reason not be able to. But it would seem that in order to discriminate any better than this, we will have to take recourse to logical criteria of some sort. Then, for example, speakers so incompetent that they can simultaneously believe the statements 'Jupiter is bigger than Mars' and 'Mars is bigger than Jupiter' could be excluded. As William Cooper (1978: 57), who has given a detailed exposition of the descriptivist view, puts it: 'If someone did claim to believe them both, one would have to challenge either his understanding of English, particularly his understanding of the full meaning of comparative construction, or else his intellectual capacity for applying his linguistic knowledge accurately in this particular situation.'

Well, take your pick. And then decide whether English speakers who simultaneously believe the statements (3) and (4) of section 1.1 should be treated in the same manner - or do they both understand the full meaning of 'if ... then' and 'not' and apply their linguistic knowledge accurately in this case?

No self respecting descriptivist will want to have anything to do with such selection procedures. Indeed, descriptivists will do their utmost to banish all logical bias, the more so as they expect their informants to do so as well. Or again, as Cooper (1978: 89) puts it, quite unaware that he might be contradicting his earlier remarks: 'In order to gain a more objective view of a language one must instead try to think like a Martian who has no idea what any of the human languages are like.'
In addition to the problem of ascertaining an informant's competence, there is also the problem of ascertaining his or her impartiality. This is no less troublesome. For one thing, it is a fact that the intuitive judgments of those who have been in contact with logical theories are in many cases different from the judgments of those who have not. Somehow a training in logic affects one's powers of judgment: once speakers have been exposed in this way their intuitions appear to be corrupted for good, no matter how they will try to unburden their mind from its theoretical load. In any case, we can never be sure if and to what extent the affected persons have retained their original powers of judgment. So, for safety's sake all those acquainted with logical theories should be excluded from having any intuitive say. Moreover, it does sometimes happen that speakers confronted with some such quirk of language as the case of the marbles quite healthily and of themselves begin to theorize, in order to remove the confusion. Their considerations may be amateurish in comparison with those of the professional logicians, but they are no less infectious. So perhaps all those suspected of theoretical tendencies should be excluded as well. But where would all of this stop and who would be left over?

The most serious limitation of this modern approach lies, however, in the fact that professionals no less than amateurs usually invent logical theories precisely where their pre-theoretical intuitions desert them - and for that reason. Aristotle's Sea Fight Argument, the Sorites Paradox, the Paradoxes of Zeno, the Liar: all contain arguments which simply are not, intuitively and without further ado, valid or invalid. The situation may be less disastrous when dealing with, say, non-referring definite descriptions, or with conditionals, but even in these cases it would be gratuitous to suppose that we all intuitively know what we are doing. It is a riddle how our pre-theoretical intuitions, vague and dubious as they often are, could serve as a touchstone in such cases.
The only way out, it seems, would be to draw a distinction between clear intuitions and the less clear, with all of the problems of operationalizing which this would bring in its train. And even if these problems, which for the concepts of impartiality and competence would seem difficult enough, can be solved, it remains to be shown that the competent and impartial speakers of a language have sufficiently many of these 'clear' intuitions to ensure that there is only one logic covering them.

The descriptivists themselves were and are among the first to recognize these difficulties with the empirical basis of what they call Natural Logic, an empirical science which aims at discovering the logic(s) underlying natural language(s). Only they do not think that these difficulties are restricted to their discipline. As a rejoinder to the criticism that the intuitions of the native speakers of a language do not constitute a rock bottom empirical basis for testing logical theories they will be inclined to invoke philosophers of science such as Lakatos and Feyerabend who said that there aren't any rock bottoms anyway - not even for most physical theories. For example, the criticism brought forward in connection with the concept of competence might be considered tantamount to the remark that a logical theory is for its testing dependent on another theory - a theory of competence. And that is common enough in the field of empirical science. As Paul Feyerabend (1970: 204) puts it: 'It is hardly ever the case that theories are directly compared with 'the facts' or with 'the evidence'. What counts and what does not count as relevant evidence usually depends on the theory as well as on other subjects which may conveniently be called 'auxiliary sciences' ('touchstone theories' is Imre Lakatos's apt expression').

Maybe this analogy is instructive and maybe other analogies can be drawn between natural logic and established empirical disciplines in order to cover some of the other difficulties
which I mentioned. But I do not think they can all be removed in this way. Take the problems I discussed in connection with the concept of impartiality. Certainly, it will not be difficult to find examples of measuring instruments which react like theoretically biased informants. To give an example, one cannot expect to falsify the statement that the volume of a fixed mass of fluid mercury at constant pressure is directly proportional to its temperature, in an experiment where the temperature is measured with the aid of a mercurial thermometer. Yet, measuring instruments which set their own standards, and even tend to change them, like informants tend to do who spontaneously start theorizing for themselves - those are unprecedented, I am afraid. Imagine a telescope lacking sufficient resolution, which when its resolution is inadequate, simply shows what it thinks might be there or should be there. If all telescopes worked that way, that would certainly cripple astronomy except if they did what they think best in a predictable manner (like the 'intelligent' television screens which increase contrast). But it is precisely this uniformity which cannot be expected in the case of self improving intuitions. People simply and as a matter of fact do not resolve the issues in the same way - otherwise there wouldn't be anything to argue about among the Natural Logicians.

I.1.3. A more pragmatic view

Suppose the logic of conditionals does exist. And suppose that we are presented with an attempt to describe it. Then the point made in the preceding pages is that we will have no way of seeing if the latter is an accurate representation of the former. At crucial points, the theoretically trained cannot be allowed to look, while the uninitiated will see only blurs.

In the pragmatic view, the logic of conditionals is not to be recognized in some logical theory; some theory is to
achieve recognition as the logic of conditionals. On this account, a theory's ratification will have little to do with its accurately predicting intuitions and everything to do with its clarifying these. The better theory is the one so well motivated that people are prepared to allow it to guide their judgments whenever their intuitions leave them groping, and even to correct their judgments whenever their intuitions do not - not yet! - match the theory's predictions.

10) Pragmatists differ from descriptivists in not assuming that there already is a logic of conditionals, much less that the competent speakers, in virtue of their competence, are implicitly acquainted with it; their confusion when faced with the eccentricities of such sentences and the many disagreements, even among specialists, on the subject of conditionals are the pragmatist's evidence for assuming that conditionals do not yet have any clear cut logic - in any case none which is accepted as such.

Now pragmatists and descriptivists differ in temperament as well. If the latter were by any chance to discover that conditionals do not yet have any logic - and surely, they must reckon with the possibility that the logic of conditionals might be simply unsettled at this point in the evolution of natural language - they would have to lay down tools. Being observers on principle, they can only wait and see if perhaps some new developments arise. Not so the pragmatists. Disregarding any advice against interfering with the natural development of language, they do approximately the following: they construct a theory which is intended as a guide to using conditionals and determine the logic conditionals would get if people were to use them the way this theory suggests.

Strangely enough, theories will not always carry the marks of their origin, whether descriptivist or pragmatist. The reason for this is that descriptivists are usually not satisfied with just predicting which arguments are valid and which are not; they want to explain things as well. In order to
achieve this, they will often incorporate into their theory a description of the way in which conditionals are used — by the competent speakers, that is. But then, of course, descriptions of competent usage can easily be interpreted as guidelines for the less accomplished, and vice versa.

This, however, should not obscure the fact that pragmatists and descriptivists take quite different actions once their theories are ready. Descriptivists make it a point of duty to test the correctness of the theory by comparing its predictions with the intuitions of the competent speakers — a fool's errand, as we saw. To pragmatists the idea of testing a theory's correctness is quite foreign. They are hardly interested whether the speakers of the language concerned already use conditionals the way they would like them to, they want to know whether the speakers are willing to do so in future. And the only way to find that out is to see how they react to the theory's explanations.

Of course, pragmatists will not be able to force us into using conditionals according to their prescriptions any more than their descriptivist colleagues can compel us to remain doing so according to their own descriptions. Ultimately, we will have to sort that out for ourselves. The difference is that whereas descriptivists must be content tagging along behind changes in usage, pragmatists take it upon themselves to bring them about: they will try to ensure that following their advice would turn language into a more useful instrument of communication.

What about the judgments of the competent and theoretically unbiased speakers of the language — don't they matter anymore? Of course they do. And so do the judgments of the less competent and the theoretically biased speakers, though on the pragmatic account a speaker's judgment on the validity of a given argument, whether intuitive or not, is in many cases just the beginning of a test rather than the end. For example, if according to the theory in question a certain
argument is logically valid whereas according to a certain informant it is not, then what matters is not so much the informant's judgment, but the considerations that have led to it. If there are no such considerations - i.e. if the informant can only take recourse to her intuitions - it may be helpful to explain how the validity comes about according to the theory in question. It might happen that after this explanation the informant changes her mind; she might even conclude that she was mistaken in her judgment. It might also occur that the informant still refuses to accept the validity of the argument. But at that stage of the discussion she will probably be able to be more explicit about her motives for doing so. The least one may expect her to do, then, is to point out what she does not like about the explanation so that one may get an indication as to why she does not want to use conditionals the way the theory suggests. It would be a lot more helpful, however, if apart from that she would offer an alternative theory and explain how her theory renders the argument invalid. Then it is the proponent of the original theory who might change his mind; he might be prepared to admit that the alternative theory suggests a better way of using conditionals than his own theory does. But it might also be that he sticks to his theory and adduces some new arguments in support of it in order to make his opponent as yet change her mind. And so on.

Admittedly, this picture lacks detail. It does not give any clue as to the kind of arguments that can play a role in the discussion. And perhaps it is too rosy a picture, too optimistic about the extent to which one or the other theory will emerge in the heat of the competition as a better theory than the others. Who is to say that the informant will allow the theoretical discussion to change her mind for her? Indeed, in this book we will on several occasions meet up with theories which are unsatisfactory because they render argument forms valid which we simply cannot get ourselves to support. The explanations which these theories give for this claim are simply not enough to bend our intuitions in their favour.
Still, the sketch given above does show how any disagreement in the validity of some concrete example naturally develops into a full-fledged theoretical discussion. And that is the point I wished to make. Theoretical discussion, something descriptivists at best condone as a marginal activity (also known as 'explaining away putative counterexamples'), is all in the day's work for a pragmaticist.

One final topic: does the pragmatic approach hold out better prospects for the problem of conditionals than the descriptivist approach? It may sound rhetorical, but this question must still be answered with some caution. Since it allows for a comparison of rival logical theories not only at the level of prediction but also at the level of explanation, the pragmatic approach enables us to evade the kind of problems that beset the descriptivist. So much, I hope, will have become clear. But new problems lie ahead. We are supposed to evaluate these theories according to the usefulness, fruitfulness, efficaciousness or what have you of the alternative ways of using conditionals which they prescribe. And at this point a good measure of scepticism may well be due. Do these notions provide any workable criteria? And if not, do they make much sense in this context?

I am afraid that the main reason why words like 'useful' and 'efficacious' slip out so easily when we are talking about ways of using 'if ... then' or other phrases, is that we are being carried away by a metaphor: words are like tools - and of course some tools are more useful than others, just as some ways of handling a given tool are more efficacious than others. In the case of real tools like hammers and saws this is fairly clear-cut, because it is obvious from the start what we want to use them for. Moreover, we can always perceive the results of applying them. Thus, we can literally demonstrate their utility. And if a certain way of handling such a tool is not very instrumental, we can always, so to speak, furnish material proof of this fact.
Now think of the results of applying a linguistic tool like 'if ... then'. Or try to explain what this tool can be used for - indeed, is 'if ... then' used for anything at all?

It is highly questionable whether the metaphor of tools puts us on the right track here. Still, the last question is meaningful, also when it is taken literally - and I remain very much inclined to answer it in the affirmative. There is more to a word than just the way in which it is used, there is also the purpose for which it is used. 11) In most cases the former is well-suited to the latter. But the very problem of conditionals is that however we use 'if ... then' (not in the least when we do so in an intuitive way), things never seem to work out quite as we want them to.

Unfortunately, the obvious next question - well then, what is the purpose of 'if ... then'; how should things work out? - is extremely difficult to answer in the abstract. Apparently, the only way to get to grips with it is by studying concrete proposals for using 'if ... then'. And even that may be too much said. The only thing that can be said for certain is that by comparing these proposals we can sometimes decide that one is more adequate to our needs than the other, without thereby being able to tell whether these needs were there from the beginning or whether they are newly arisen ones, aroused as it were by the comparison itself.

So the criterion of usefulness, even though it is ultimately decisive, is not very workable in practice. One cannot make up, in advance, a list of requirements that any useful way of using 'if ... then' should meet - no definitive list, that is. Besides, the relative usefulness of any proposed way of using 'if ... then' is certainly not the only thing that matters. For one thing, it might very well be that some of these proposals do not provide anything worth calling a way of using 'if ... then' (or any other expression for that matter) in the first place - not so at least in the eyes of someone whose own proposal is based on
an altogether different philosophy of language. To put it otherwise: the question which theory suggests the most useful way of using 'if ... then' only becomes pertinent when we are comparing theories developed within the same theoretical framework. When the frameworks differ more general questions will arise, pertaining to the frameworks themselves rather than to the particular theories developed within them.

Since the problem of conditionals has prompted several alternative approaches to logic as a whole - not even the concept of validity has been kept out of harm's way - we shall more than once have an opportunity to discuss such general questions in the sequel. It will appear that not all of them can be decided on purely pragmatic grounds. Some will be rather metaphysical in character, other ones will concern epistemological matters. What I have called a theoretical framework contains among other things a description of the kind of circumstances (ontological and epistemological) in which the language users find themselves - in which, so to speak, they cannot help but find themselves. This kind of circumstances in its turn puts heavy constraints on the possible ways in which they might use 'if ... then'. Therefore, if we ever want to devise a way of using 'if ... then' that is of any use at all we'd better reckon with the circumstances as they really are.

The predictions of a logical theory are of minor importance, what matters is its explanatory force. For a logical theory has to be sold rather than tested. None of the comments made above should be allowed to obscure this fact. At best they tell that if we really want our theories to be bought, we must take care that they satisfy the actual needs of the language users. Not anything goes; there are all kinds of reasons why people might not be willing to buy a theory. Nevertheless, sometimes they do. Sometimes the way of speaking suggested by some logical theory is recognized by a
large group of people as a useful way of speaking. Take classical first-order logic for example. Admittedly, the way of speaking underlying this logic does not serve all purposes equally well. Accordingly, the campaign to sell it as such, as was once the intention of Ideal Language Philosophy, has failed. Still, the way of speaking that goes with classical first-order logic has been recognized as most useful for mathematical purposes, in particular by those philosophers of mathematics who take a realist stand on the ontological status of mathematical objects. They made classical first-order logic the underlying logic of set theory. And now that set theory has become widely accepted as the basis of all mathematics, all students of mathematics are taught to express themselves as first-order logic says they should. In this book, too, especially in the more technical parts of it, classical first-order logic, together with the 'material' way of using 'if ... then' that goes with it, is adopted as the standard of reasoning for the metalanguage. 12)
I.2. EXPLANATORY STRATEGIES

The leading theme of the previous chapter may not have made much of an impression on any who have themselves not experienced how deeply theoretical arguments can affect one's intuitions. Hopefully the following pages, which contain an extensive discussion of the case of the marbles will strengthen my point.

My principal concern, however, will be to introduce the various schools of thought in the field of conditional logic and to discuss the kinds of arguments which they employ. The marbles puzzle is not discussed for its own sake, but to this end.
I.2.1. Logical validity

I.2.1.1. The standard explanation

The usual explication of logical validity runs as follows:

(S1) An argument is logically valid iff its premises cannot all be true without its conclusion being true as well

By far the most theories of conditionals developed so far start from this explanation of logical validity; and nearly all of these are based on the additional presumption that non-truth equals falsity:

(S2) No sentence is both true and false
(S3) Every sentence is either true or false

Granted (S2) and (S3) the following truth-condition for negative sentences needs no further comment.

(S~) A sentence of the form $\neg \phi$ is true iff $\phi$ itself is false

Once you have come to accept all this, you hardly need a theory of conditionals to see that the Principle of Modus Tollens is logically valid. All you need is this sufficient condition for the falsity of a conditional sentence.

(SF) Any sentence of the form $\phi \rightarrow \psi$ is false if its antecedent $\phi$ is true and its consequent $\psi$ is false

All theories based on (S1), (S2) and (S3) not only subscribe to (SF) but also to the following sufficient condition for the truth of a conditional sentence.

(ST) Any sentence of the form $\phi \rightarrow \psi$ is true if its consequent $\psi$ logically follows from its antecedent $\phi$
I.1. PROPOSITION. Given \((S_1), (S_2), (S_3), (S\sim),\) and \((SF)\) the principle of Modus Tollens is logically valid.

PROOF. Suppose Modus Tollens is not valid in the sense of \((S_1)\). Then it should be possible for there to be three sentences of the form \(\phi \rightarrow \psi, \sim \psi\) and \(\sim \phi\) such that both the first and the second are true and the third is not true.

If \(\sim \phi\) is not true, then by \((S\sim)\) \(\phi\) is not false; and if \(\phi\) is not false, then by \((S_3)\) \(\phi\) is true. If \(\sim \psi\) is true, then by \((S\sim)\) \(\psi\) is false. So \(\phi\) is true and \(\psi\) is false, which with \((SF)\) yields that \(\phi \rightarrow \psi\) is false. Given \((S_2)\), this contradicts the requirement that \(\phi \rightarrow \psi\) should be true. \(\Box\)

All theories based on \((S_1), (S_2)\) and \((S_3)\) subscribe to \((S\sim)\) and \((SF)\). In other words, all of them sanctify Modus Tollens. Consequently their advocates will maintain that it is wrong to accept both the sentences \((3)\) and \((4)\) — or at any rate their formal counterparts \((3')\) and \((4')\) — occurring in the marbles puzzle. At least one of these premises must be rejected, but which one should it be? As I noted before, the consensus seems to dissolve here just as rapidly as it was reached. I shall present two theories, the one requiring us to accept \((4')\) and not to accept \((3')\), the other one precisely the opposite.

The first theory tells us that 'if ... then' is just the so-called material implication: the condition laid down in \((SF)\) is not only sufficient but also necessary for the falsity of conditional sentences. Using \((S_2)\) and \((S_3)\) this means:

\[(\Rightarrow) \text{ A sentence of the form } \phi \rightarrow \psi \text{ is true iff its antecedent } \phi \text{ is false or its consequent } \psi \text{ is true}\]

This has certainly been the most disputed truth-condition in the history of logic ever since the Megarian Philo (fourth century B.C.) first suggested it — not only the most heavily criticised, but also the most ably defended. In recent
introductory textbooks nobody ever pretends that (\(\triangleright\)) exhausts the meaning of 'if ... then'. At best it is pointed out that it is the only alternative left once one has chosen for (S2) and (S3), assuming one insists on speaking a *truth-functional* language, i.e. a language for which the following holds:

\[ (T) \text{ The truth value of any compound sentence is uniquely determined by the truth values of its constituent sentences} \]

Given (S2), (S3) and (T), the question as to which truth condition is best for conditional statements boils down to the question as to which of the values 'true' or 'false' the compound sentence \(\phi \rightarrow \psi\) should be assigned in each of the following cases: (i) \(\phi\) is true, \(\psi\) is true; (ii) \(\phi\) is true, \(\psi\) is false; (iii) \(\phi\) is false, \(\psi\) is true; (iv) \(\phi\) is false, \(\psi\) is false. The answer is: (i) true; (ii) false; (iii) true; (iv) true. (Proof: consider the following conditional sentence *If Allen is over fifty years old, then he is over thirty years old*. We certainly want this sentence to come out 'true', whatever Allen's age may in fact be. Now suppose Allen is in fact sixty; then the antecedent is true, and so is the consequent (case (i)). If Allen is in fact forty, the antecedent is false and the consequent true (case (iii)). And if he turns out to be twenty, both the antecedent and the consequent are false. So there is at least one example of a conditional sentence that is true in case (i), true in case (iii) and true in case (iv). But if one conditional sentence is true in these cases, then by (T) all are. Hence, we must put the value 'true' in case (i), (iii) and (iv). That we must put the value 'false' in case (ii) is now obvious if only because otherwise every conditional sentence would turn out to be a logical truth.)

The logical theory of truth-functional languages is relatively simple. Therefore from a didactic point of view it is preferable to discuss the properties of those languages first in an introductory course. But does (T) have any merits
over and above this one? Why would anyone want to speak a
truth-functional language? The answer to this question is,
I think, to be found not so much in (T) alone, but in the
combination of (T) and

(A) The truth value of any non-compound sentence is solely and
entirely dependent on what is, in fact, the case

(T) and (A) are two of the cornerstones of Wittgenstein's
Tractatus Logico-Philosophicus. In a way, they just restate the
old positivist principle that only facts can be a genuine
source of knowledge: the truth value of any statement is
wholly dependent on what is, in fact, the case. But (T) and
(A) do more than just restate this principle, they also tell
how one might live up to it: speak a truth-functional
language, and things will work out exactly as you want them
to. No wonder, then, that Wittgenstein's Tractatus found so
much response among the members of the Vienna Circle.

We are ready now to apply this theory to the marbles
puzzle. Applying (⇒) to (4'), we see that it is false just in
case all three marbles are in box 2. But we already know that
this cannot possibly be so. Only the situations depicted in
the table can obtain, and whichever of these may happen to be
the real situation, (4') will hold. Therefore, we can safely
accept (4').

It is on the other hand not possible to say anything
definite about the truth value of (3'). (⇒) and (S~) say that
(3') is true iff both blue and yellow are in box 2. So
(3') is true if the real situation is like the one
depicted in III, but false if the real situation happens to
be any of the others. Hence, it would be premature to say
anything definite about (3') at this stage. We cannot accept
it, but we cannot reject it either. This is not to say (3')
has not got any definite truth value, only that we lack the
information to decide which truth value.
The second theory I want to discuss would have us believe that 'if ... then' is a so-called strict implication. Roughly speaking, a conditional sentence is false according to this theory not only if it is in fact the case that its antecedent is true and its consequent false, but also if that might possibly be so. Just to get an idea of how this works, we turn to (3') again. Clearly, the blue marble might be in box 2 together with the yellow one (witness situation III). This means that the sentence \( \text{blue in } 2 \rightarrow \text{yellow in } 1 \) evaluated as a strict conditional, turns out false. Consequently its negation (3') turns out true.

Before we can say what happens to (4') we must develop the rough idea given above more fully. Note first that the principle (T) of truth-functionality is abandoned by this theory: the truth value of a conditional is not uniquely determined by the truth values of its constituents - not by their actual truth values at least. In evaluating a conditional we must reckon not only with the truth values which its constituents happen, in fact, to have, but also with the truth values which these might possibly have. We are as it were to transfer ourselves to situations other than the actual one, and to evaluate the constituents there. If we are dealing with constituents that are themselves conditionals then evaluating the constituents in these other situations will involve transferring ourselves to yet 'other' other situations, and so on.

One way\(^{13} \) of working this out is to say

\[ (\Rightarrow) \text{ A sentence of the form } \phi \rightarrow \psi \text{ is true in a given possible situation } s \text{ iff there is no possible situation } s' \text{ such that } \phi \text{ is true in } s' \text{ and } \psi \text{ is false in } s' \]

Truth simpliciter is then to be understood as truth in the actual situation.

Now, as far as the case of the marbles is concerned, the relevant possible situations are the ones depicted in the table.\(^{14} \) There is one situation, III, in which the blue and the yellow marble are both in box 2. So, by (\( \Rightarrow \)) we have that
the sentence \((\text{blue in 2} \rightarrow \text{yellow in 1})\) is false in each of the situations I-VI. In some of these, notably I, II and VI, the red marble is in box 2. So there are situations in which the sentence \(\text{red in 2}\) is true while the sentence \((\text{blue in 2} \rightarrow \text{yellow in 1})\) is false. Given \((\neg\neg)\), this means that \((4')\), \(\text{red in 2} \rightarrow (\text{blue in 2} \rightarrow \text{yellow in 1})\) is false in each of the situations I-VI. In particular we have that \((4')\) is false in the actual situation, whichever situation that may be. Therefore, it is wrong to accept \((4')\).

It is of interest to compare \((4')\) with

\[(1') \ (\text{red in 2} \land \text{blue in 2}) \rightarrow \text{yellow in 1}\]

Given the following truth condition for conjunctions it can easily be verified that unlike \((4')\), \((1')\) is true.

\[(S\wedge) \ A \text{ sentence of the form } \phi \land \psi \text{ is true in a given possible situation } s \text{ iff both } \phi \text{ and } \psi \text{ are true in } s\]

Apparently the theory of strict implication distinguishes between sentences of the form \((\phi \land \psi) \rightarrow \chi\) and sentences of the form \(\phi \rightarrow (\psi \rightarrow \chi)\). The theory of material implication does not, and - intuitively - that is a point in its favour. But the material implication has many problems of its own, as we will see in due course. So the question arises whether or not there are any other theories in the framework described by \((S1)\), \((S2)\), \((S3)\), \((S\neg)\), \((S\wedge)\), \((SF)\) and \((ST)\) which render these two argument forms equivalent.

Its answer\(^15\) is that there are not.

I.2. PROPOSITION. Within the framework given by \((S1)\), \((S2)\), \((S3)\), \((S\neg)\), \((S\wedge)\), \((SF)\) and \((ST)\), the theory of material implication is the only theory which renders the argument form

\[(* \ (\phi \land \psi) \rightarrow \chi \ / \ \phi \rightarrow (\psi \rightarrow \chi)\]

valid.
REMARK. Here and elsewhere I write \( \phi_1, \ldots, \phi_n \vdash \psi \) for the argument with the set of premises \( \{\phi_1, \ldots, \phi_n\} \) and the conclusion \( \psi \).

PROOF. We must show that the only truth condition for sentences of the form \( \phi \rightarrow \psi \) left over is this:

\( \phi \rightarrow \psi \) is true iff \( \phi \) is false or \( \psi \) is true

The proof from left to right has already been given in the proof of proposition 1.

For the converse, we must show that

(i) if \( \phi \) is false, then \( \phi \rightarrow \psi \) is true;
(ii) if \( \psi \) is true, then \( \phi \rightarrow \psi \) is true.

(i) Suppose \( \phi \) is false. By (S~), it follows that \( \sim \phi \) is true

Note that given (S1), (S2), (S~), and (S~), \( (\sim \phi \land \phi) \vdash \chi \) is valid. In view of (ST) this means that \( (\sim \phi \land \phi) \vdash \chi \) is true.

Using (*) we see that \( \sim \phi \rightarrow (\phi \rightarrow \chi) \) is true. Since \( \sim \phi \) is true, it follows that \( (\phi \rightarrow \psi) \) is not false, otherwise (SF) would not hold. So, by (S2) \( \phi \rightarrow \psi \) is true.

(ii) Analogous. (Note that given (S1) and (S~) \( \psi \land \phi \vdash \psi \) is valid.)

I.2.1.2. Truth and evidence

In part II of this book we will return at some length to the standard concept of validity and the various theories of conditionals founded on it. In part III, a theory is developed on the basis of a different explication of validity. Compare (D1) with (S1):

\[ (D1) \text{ An argument is logically valid iff its premises cannot all be true on the basis of the available evidence without its conclusion being true on the basis of that evidence as well} \]

We obtain (D2), (D~) and (DF) in exactly the same way from (S2), (S~) and (SF) by substituting 'true on the basis of the available evidence' and 'false on the basis of the available evidence' for 'true' and 'false' in the original.
We cannot translate (S3) in this manner, since the result of doing so is unacceptable: not every sentence need to be decided, true or false, by the available evidence. For example, the data presently at your disposal neither allow you to conclude that the red marble is in box 2, nor that it is not.

(DF) offers only a sufficient condition for a sentence of the form \( \phi \rightarrow \psi \) to be false on the basis of the available evidence. As a first approximation of a necessary and sufficient condition we have

\[ (D \rightarrow) \text{ A sentence of the form } \phi \rightarrow \psi \text{ is false on the basis of the available evidence iff this evidence could develop into evidence on the basis of which } \phi \text{ is true and } \psi \text{ is false. Otherwise, } \phi \rightarrow \psi \text{ is true on the basis of the available evidence.} \]

Let us apply these ideas to the marbles puzzle, beginning with \((3')\). Clearly we could at some later stage be less ignorant about the exact distribution of the marbles among the boxes. We might for example find out that the distribution in III of the table is the real one. There both the blue and the yellow marble are in box 2. This means that the evidence we have could develop into evidence on the basis of which the sentence \textit{blue in 2} is true and the sentence \textit{yellow in 1} is false. So, by \((D \rightarrow)\), \textit{blue in 1} \rightarrow \textit{yellow in 1} is false on the basis of the evidence presently available and with \((D \sim)\) this gives that its negation \(\neg(\text{blue in 2} \rightarrow \text{yellow in 1})\) is true.

As far as \((4')\) is concerned, suppose that our present information about the distribution of the marbles was to grow in such a way that we knew its antecedent, \textit{red in 2}, to be true. Then \((4')\)'s consequent \textit{blue in 2} \rightarrow \textit{yellow in 1} would also be true on the basis of the available evidence. This is easily seen as follows. If the information could grow some more in such a way that we were to learn that the blue marble is in box 2 and the yellow is not in box 1, we would
at that stage have evidence on the basis of which all three sentences 'red in 2', 'blue in 2', 'yellow in 2' were true. But our initial information that there is at least one marble in each box excludes this possibility.

According to this theory, then, both (3') and (4') are quite acceptable, while Modus Tollens fails. It does not follow from (3') and (4') that the red marble is not in box 2; we only get that it may not be there.

Still, this theory does not leave us very much room to tinker with Modus Tollens. (D1), (D2), (D~) and (D+) preclude three sentences of the form \( \sim \psi \), \( \phi \to \psi \) and \( \phi \) all being true on the basis of the same evidence. So, what happens if we are finished with the extra information that the red marble is in box 2? Well, hopefully something will happen to the truth of (3') or (4'). It does, as can easily be verified. On the basis of such new information, (3') will be false. What this means is that truth or falsity on the basis of the available evidence need not be invariable under growth of that evidence. Various sorts of sentences do possess this sort of stability, but conditional sentences typically do not. A conditional \( \phi \to \psi \) can be false on the available evidence simply because it is not yet possible to rule out the possibility that \( \phi \) will turn out true without \( \psi \) turning out true too: the available evidence is just too scanty. Adding information which does rule this possibility out will then switch the truth value of the conditional.

It is worth noticing that this approach in a sense lies somewhere between conditionals as material implications and conditionals as strict implications. In evaluating a material implication, one is only interested in what holds in reality - one behaves, so to speak, as if the evidence is complete. In evaluating a strict implication, one always takes all possibilities into account - as if, so to speak, one never learns that some of these are, in fact, excluded. But according to this approach, one is interested only in those possibilities which are left open by the evidence which one happens to have. This, of course, makes
conditionals highly context dependent. In a context where no specific evidence is available, they are like strict implications. In a context where the evidence is complete, they are like material implications. In most contexts they are neither.

I.2.1.3. Probability semantics

The theory developed in part III is not the first theory of conditionals based on a non-standard explication of logical validity. Indeed, there is something about conditionals which seems to invite such a manoeuvre.

Ernest Adams (1966, 1975) has also proposed a modification of the classical standard of logical validity. His idea is that some conclusion logically follows from a set of premises not if its truth is guaranteed by theirs, as the classical standard would have it, but rather its probability. An argument is said to be valid, on Adam's view, if it is possible to bring the probability of its conclusion arbitrarily close to one by raising that of its premises above some suitable value. More precisely,

(P1) An argument \( \Delta / \phi \) is valid if for every \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that for any probability assignment \( P \) if \( P(\psi) > 1-\delta \) for each \( \psi \in \Delta \), then \( P(\phi) > 1-\varepsilon \).

Consider the set \( L_0 \) of sentences that can be formed out of the atomic red in 1, blue in 1, yellow in 1, red in 2, blue in 2, and yellow in 2 by conjunction and negation only. The next definition says what a probability assignment to the sentences of \( L_0 \) is.

(P2) A probability assignment for \( L_0 \) is a function that assigns a value \( P(\phi) \), 0 \( \leq P(\phi) \) \( \leq 1 \), to every \( \phi \in L_0 \), while furthermore

(i) \( P(\phi) = P(\phi \land \phi) \)

(ii) \( P(\phi \land \psi) = P(\psi \land \phi) \)

(iii) \( P(\phi \land (\psi \land \chi)) = P((\phi \land \psi) \land \chi) \)
\{i\} P(\phi) + P(\neg\phi) = 1
\{ii\} P(\phi) = P(\phi \land \psi) + P(\phi \land \neg\psi)

Think of the probability P(\phi) of a sentence \phi as an agent's degree of belief in \phi - something that could be measured by betting odds: if you are willing to bet at odds of 1:5 - but at no higher odds - say 1:1 - for the proposition that both the red and the blue marble are in box 2, then your degree of belief P(red in 2 \land blue in 2) = 1/1+5 = 1/6 - rather than 1/1+1 = 1/2.

(P2) imposes heavy constraints on the way an agent can distribute his degrees of belief among the various sentences of L. Actually, it is not meant as a description of how a particular agent will do so, but of how a rational agent would. Some distributions of belief are not rational in that they allow a Dutch book to be made against anyone whose state of beliefs conforms to them. (A Dutch book is a set of bets which the holder of the beliefs must accept (given his belief distribution) but which will certainly result in a net loss to him in the long run.) It can be shown that a necessary and sufficient condition of not being in a position to have a book made against you is that your degrees of belief in the sentences of L_0 satisfy the axioms laid down in (P2), which are just a version of the axioms of the probability calculus.

Example: Assuming that you find each of the six distributions of the marbles among the two boxes equally probable (and assuming you do not consider any other distribution possible), you should get

P(red in 1) = 1/2
P(\neg(red in 2 \land blue in 2)) = 5/6

Now, what is the logic generated by (P1) and (P2)?
I.3. PROPOSITION. Let $\Delta/\phi$ be an argument of $L_0$.
$\Delta/\phi$ is valid according to (P1) and (P2) if and only if
$\Delta/\phi$ is valid according to (S1), (S2), (S3), (S∼) and (S∧).

PROOF. For a proof of this proposition the reader is referred to Adams (1975: 57-58).

In other words, as far as $L_0$ is concerned, the probabilistic account does not bring up anything new. Things change, however, if we turn to sentences containing 'if'.
The basis for the probabilistic treatment of the conditional sentences consists, not surprisingly, in the idea of conditional probability.

(P3) Let $P$ be a probability assignment to the sentences of $L_0$
and $\phi$ a sentence of $L_0$ with $P(\phi) \neq 0$. We define $P_\phi$ as the function assigning to every sentence $\psi$ of $L_0$ the value
$$P_\phi(\psi) = \frac{P(\phi \land \psi)}{P(\phi)}$$

I.4. PROPOSITION. Let $P$ and $P_\phi$ be as above.

$P_\phi$ is a probability assignment in the sense of (P2).

PROOF. Left to the reader.

$P_\phi(\psi)$ is called the conditional probability of $\psi$, given $\phi$.
Intuitively what $P_\phi$ is supposed to describe is the change of belief that results once $\phi$ is known to be true. Again it can be proved that agents who do not change their beliefs by 'conditionalizing' on this new information can always have a Dutch book made against them.

Example: Assuming, once again, that your degree of belief in each of the six possible distributions of the marbles among the two boxes is 1/6, you should get

$$P_{\text{blue in 2}}(\text{yellow in 1}) = \frac{2}{3}$$

$$P_{\text{red in 2}} \land P_{\text{blue in 2}}(\text{yellow in 1}) = 1$$

$$P_{\text{red in 2}} \land P_{\text{blue in 2}}(\text{yellow in 1}) = 1$$
Notice that $P_{\phi \psi}(\chi) = P_{\phi \land \psi}(\chi)$ for any $\phi, \psi, \chi$ such that $P(\phi \land \psi) \neq 0$.

Now let us see how we must adapt (P2) if $L_0$ is extended to $L$, the set of sentences that can be built from atoms red in 1, red in 2 etc. using conjunction, negation and implication (and nothing else).

The basic idea is to equate the absolute probability of the conditional $\phi \rightarrow \psi$ with the conditional probability of $\psi$, given $\phi$.

(P4) If $P(\phi) \neq 0$, then $P(\phi \rightarrow \psi) = P_{\phi}(\psi)$

If $P(\phi) = 0$ we leave $P(\phi \rightarrow \psi)$ undefined. (This means of course that we will have to reformulate the axioms given in (P2) in such a way that they do not apply to sentences whose probabilities are undefined.)

As an example of this idea working at its best, consider the Hypothetical Syllogism

$\psi \rightarrow \chi$, $\phi \rightarrow \psi / \phi \rightarrow \chi$

The relevant probabilities are then the following

\[
P(\psi \rightarrow \chi) = \frac{P(\psi \land \chi)}{P(\psi)},
\]

\[
P(\phi \rightarrow \psi) = \frac{P(\phi \land \psi)}{P(\phi)}, \text{ and}
\]

\[
P(\phi \rightarrow \chi) = \frac{P(\phi \land \chi)}{P(\phi)}
\]

We can choose mutually exclusive $\phi$ and $\chi$, and a sentence $\psi$ which is compatible with both of these. Then $P(\phi \rightarrow \chi)$ will be zero. But the probabilities of $P(\psi \rightarrow \chi)$ and $P(\phi \rightarrow \psi)$ can increase without this having any effect on $P(\phi \rightarrow \chi)$. So the Hypothetical Syllogism is not in general valid in the sense of (P1) which is surprising but just as well. No one would accept that from
If Jones wins the election then Smith will retire to private life and

If Smith dies before the election then Jones will win it

it follows that

If Smith dies before the election he will retire to private life

(P4), however, cannot be all there is to it. For one thing, it does not allow us to extend the result mentioned in proposition 4 to the language $L$. That is, we cannot be sure that we still get an (extended) probability assignment after conditionalizing on an (extended) probability assignment.

It seems obvious that the way to achieve this would be to add the following requirement:

(P5) If $P(\phi \land \psi) \neq 0$ then $P_\phi(\psi \rightarrow \chi) = P_{\phi \land \psi}(\chi)$

Unfortunately, this does not work. As David Lewis (1976) showed, this way you end up with probability functions which are at most four valued.

I.5. PROPOSITION. Let $P$ be an (extended) probability assignment. Suppose there are sentences $\phi$ and $\psi$ such that both $P(\phi \land \psi) > 0$ and $P(\phi \land \neg \psi) > 0$. Then $P_\phi(\psi) = P(\psi)$.

PROOF. Note first that $P(\phi) > 0$, $P(\psi) > 0$, and $P(\neg \psi) > 0$. Therefore the following makes sense.

By (P5) and (P3) we have

$$P_{\psi}(\phi \rightarrow \psi) = P_{\phi \land \psi}(\psi) = \frac{P(\phi \land \psi \land \psi)}{P(\phi \land \psi)} = 1$$

$$P_{\neg \psi}(\phi \rightarrow \psi) = P_{\phi \land \neg \psi}(\psi) = \frac{P(\phi \land \neg \psi \land \psi)}{P(\phi \land \neg \psi)} = 0$$

Furthermore, by (v) of (P2), (P3) and (P4),

$$P_\phi(\psi) = P(\phi \rightarrow \psi)$$
$$= P_{\psi}(\phi \rightarrow \psi) \cdot P(\psi) + P_{\neg \psi}(\phi \rightarrow \psi) \cdot P(\neg \psi)$$
$$= .1 \cdot P(\psi) + 0 \cdot P(\neg \psi)$$
$$= P(\psi)$$
I happen to believe that $P(\text{red in } I \land \text{blue in } I) = 1/6$, and that $P(\text{red in } I \land \neg\text{blue in } I) = 1/3$. Furthermore I would say that $P_{\text{red in } I}(\text{blue in } I) = 1/3$, and that $P(\text{blue in } I) = 1/2$. But the above forbids this. Lewis (1976) puts it more generally: '... if we take three pairwise incompatible sentences $\phi$, $\psi$ and $\chi$ [I replace Lewis's notation by mine here, F.V.] such that $P(\phi)$, $P(\psi)$ and $P(\chi)$ are all positive and if we take $\Theta$ as the disjunction $\phi \lor \psi$, then $P(\Theta \land \phi)$ and $P(\Theta \land \neg\psi)$ are positive but $P_{\Theta}(\psi)$ and $P(\psi)$ are unequal. So there are no such three sentences. Further, $P$ has at most four different values. Else there would be two different values of $P$, $x$ and $y$, strictly intermediate between 0 and 1 such that $x + y \neq 1$. But then if $P(\phi) = x$ and $P(\psi) = y$ it follows that at least three of $P(\phi \land \psi)$, $P(\neg\phi \land \psi)$, $P(\phi \land \neg\psi)$, and $P(\neg\phi \land \neg\psi)$ are positive, which we have seen impossible.

The reaction of Adams to this perplexing triviality result is to shrink the domain of application of his theory such that the above argument cannot be set up. Starting from the idea that an assertion of a conditional is a conditional assertion and that as such conditionals lack the truth-values of ordinary assertions, he argues that condition (P4) only holds for conditionals $\phi \rightarrow \psi$ whose antecedent $\phi$ and consequent $\psi$ do not contain other conditionals, and that it is wrong to ask for a generalization to other cases. Only unconditional consequents can be asserted conditionally and that only on non-conditional conditions. He even denies that one can attach probabilities to conjunctions, negations and other truth-functional compounds of conditionals. 17)

As a consequence Adams' theory cannot help us solve the puzzle of the marbles. Both (3') and (4') fall outside the scope of his theory.

Lewis' triviality result does not only pose a problem for Adams, but for everyone who wants to attach probabilities to conditionals. In what follows I will not have much to say on this problem. 18) Suffice it to say that from the
perspective of data semantics it seems misguided to try attaching probabilities to conditionals. Roughly the argument is this: bets can only be laid on sentences that are stable in the sense that once they have turned out to be true/false on the basis of the available evidence, they remain so. In the preceding section we saw that conditionals do not have this property. Now, at which point will it be decided who has won?

I.2.1.4. Relevance logic

Another and quite different criticism of the standard notion of logical validity is to be gleaned from the work of the relevance logicians. They believe that for an argument to be valid, it is not sufficient that the truth of the premises be transferred to the conclusion. The premises of the argument must in addition be relevant to its conclusion. There is something in this. It is at least misleading to conclude from the irrelevant coincidence of it raining in Ipanema that the red marble either is or is not to be found in box 1. As Anderson and Belnap (1975: 14) put it:

'Saying that $\psi$ is true on the irrelevant assumption that $\phi$ is not to deduce $\psi$ from $\phi$, nor to establish that $\phi$ implies $\psi$ in any sensible sense of implies. Of course we can say Assume that snow is puce. Seven is a prime number. But if we say Assume that snow is puce. It follows that (or consequently, or therefore, or it may validly be inferred that) seven is a prime number, then we have spoken falsely.'

Under this banner they embarked on the ambitious programme of analyzing the relation of entailment in such a way as to circumvent these and other 'fallacies of relevance'.

The explication of logical validity developed in part III is not going to satisfy the relevance logicians any more than the classical one, and for the same reasons. It does not take any account of the relevance of the premises of an argument to its conclusion in assessing its validity. And this is not the only fault which they will find.
We turn to the marbles for an example. The argument

The red marble is in box 1
If it is raining in Ipanema, then the red marble is in box 1

is valid according to standards set in part III. The problem which relevance logic would have with this is not that its premise is irrelevant to its conclusion, but rather that the antecedent of the latter is irrelevant to its consequent. (As a matter of fact, relevance logicians scarcely distinguish these two levels.) As a result, the conclusion will be deemed false and the argument form will be deemed invalid.

Everyone would agree that to argue irrelevantly is to argue badly, as it is to argue from false premises, in a roundabout way, or to the wrong conclusion. And the claim that the antecedents of conditionals should be relevant to their consequents also has something to say for it. In any case, as anyone who has ever taught truth tables knows, this idea appeals to a wider group than just the relevance logicians (compare this with section III.2), and it is to their credit that they have insisted that these matters should not be forgotten.

It seems to me, however, that the difference between finding an argument invalid because the premises are irrelevant to the conclusion, and finding it valid though ineffective for the same reasons, is largely a verbal one. Besides this, from a methodological point of view, it is dubious whether there are any advantages in lumping together these various ways in which arguments can be improper. The relevance logicians run the risk of turning logical validity into a clumsy thing. The difficulties they have in providing their largely proof-theoretic theories with a proper semantics may be regarded as a symptom of this. The semantic theories which have thus far been put forward tend to lack
the explanatory power which is to be expected from theories which purport to say what relevance means. They are in a sense merely formal, and are extremely difficult to apply in analyzing the sorts of things which we are interested in here. This applies not only to the larger part of the work done in this tradition, which is primarily concerned with the abstract notions of relevance and entailment, but also to the work done by Barker (1969) and Bacon (1971), which does focus attention on conditional sentences derived from natural language. In trying to apply their ideas to the marbles puzzle, for example, I have not been able to decide which of (3') and (4') they would recommend rejecting. (They must reject at least one of the two, since Modus Tollens is valid in relevance logic.)

I.2.2. Pragmatic correctness

One cannot assert any sentence at any time; statements can be conversationally out of place even though they are true, highly probable, or true on the basis of the available evidence. Having made this trifling observation — after all $7 + 5 = 12$ — we might ask for the criteria by which it can be determined whether or not a statement is conversationally correct. Following Grice (1975) we might try to find these criteria in some maxims of conversation which the participants in a conversation should observe in order that their conversation be as productive as possible. Then, having found these criteria, we might carry on and try to map out the circumstances under which various kinds of statements can properly be made. In doing so we would discover that statements when used correctly convey much more information than just their logical content. In addition to this there are also the pragmatic implicatures. For example, we might find that an indicative conditional if it is the case that $\phi$ then it is the case that $\psi$ usually 'implicates' both it may be the case that $\phi$ and it may be the case that not $\phi$ — usually, but not always:
implicatures are cancellable. When one asserts

if she is under twenty, then I'll eat my hat

one is not intending to suggest that she may be under
twenty at all. And the statement

she is on the wrong side of thirty, if she is a day

does not implicate that she may be less than a day old.

We would also discover that sometimes it is very hard to
distinguish between logical consequences and pragmatic
implicatures; especially if we are not yet completely sure
which logic we are dealing with. Is $\phi + \psi$ a logical or a
pragmatic consequence of $\neg \phi \lor \psi$? If you think that $+$ behaves
as a material implication, then undoubtedly you will say
that it is the first. But if like Adams you believe that $+$
behaves as a conditional probability, in which case you'll
find the inference pattern $\neg \phi \lor \psi / \phi + \psi$ logically invalid,
you will agree that it is the latter. As Adams (1966: 285)
puts it:

'What the present theory shows is that inferring $\text{if } \phi$
then $\psi$ from either not $\phi$ or $\psi$ is not always reasonable, but that
the only situation under with either not $\phi$ or $\psi$ has a high
probability, but $\text{if } \phi$ then $\psi$ has a low one is the situation
in which not $\phi$ has a high probability. Assuming this, we
have an immediate explanation of why we are ordinarily
willing to infer $\text{if } \phi$ then $\psi$ from either not $\phi$ or $\psi$: the reason
is that people do not ordinarily assert a disjunction when
they are in a position to assert one of its members
outright. (In fact, it is misleading to do so, and
therefore doing it probably runs against strong conventions
for the proper use of language.) Thus, if one heard it said
that either the game will not be played tomorrow, or the Dodgers will
win he would be well justified in inferring $\text{if the game is}
played tomorrow, then the Dodgers will win, and what would justify
the inference would be the knowledge that the person
asserting either the game will not be played or the Dodgers will win
did not do so simply on the grounds of the information he
had to the effect that the game would not be played'.
Here we see pragmatic considerations being invoked to explain why a certain logically invalid inference pattern has so many intuitively sound instances. And below we see David Lewis (1976: 137) invoking exactly the same pragmatic considerations in order to explain why a certain logically valid inference pattern has so many intuitive counterexamples. It concerns the scheme \( \lnot \phi / \phi \rightarrow \psi \).

'The speaker ought not to assert the conditional if he believes it to be true predominantly because he believes its antecedent to be false, so that its probability of truth consists mostly of its probability of vacuous truth. It is pointless to do so. And if it is pointless, then also it is worse than pointless: it is misleading. The hearer, trusting the speaker not to assert pointlessly, will assume that he has not done so. The hearer may then wrongly infer that the speaker had additional reason to believe that the conditional is true, over and above his disbelief in the antecedent.'

As these examples illustrate, Gricean arguments have become standard repertoire in defending logical theories. The idea of a pragmatic theory complementing a semantic theory has become quite familiar. Now, everyone is in agreement that semantics and pragmatics should cooperate in this way, but there is a lot less agreement as to the distribution of labour among the two. The problem is that it is hard to say where semantics stops and pragmatics takes over. What the one author classifies as a clearcut counterexample to a putative logical principle is for the other merely an innocent pragmatic exception to an otherwise faultless logical rule.

Are there any general criteria which can be used to decide who is right and who is wrong, to distinguish the domains of semantics and pragmatics? I have come to believe that only global criteria would be of any use in this, that is doesn't make much sense to take any particular inference rule and to fight it out over that one isolated example. Instead, what matters is the way a combination of a logic and its complementary pragmatic theory performs in
general. The best combination will be something like the combination which best explains the plausibility of as many plausible sounding inferences as possible, and best explains the implausibility of the rest. And the best dividing line between semantics and pragmatics will be the line drawn by whichever combination this is.

Of course this is both a simplified and rosy view of the matter. In practice, we do not have dividing lines, but gaps. Take for example classical logic. There are plenty of clearcut intuitive counterexamples to the classically valid scheme \( \neg (\phi \rightarrow \psi) \lor \phi \). For instance,

\[
\text{It is not the case that if the peace treaty is signed, war will be prevented.}
\]

The peace treaty will be signed

As yet, however, no adequate pragmatic explanation in terms of maxims being overruled by anyone arguing in this manner has been provided. As a second example, take the classically valid scheme \( (\phi \land \psi) \rightarrow \chi \lor (\phi \rightarrow \chi) \lor (\psi \rightarrow \chi) \) and its following instantiation:

\[
\text{If both the mainswitch and the auxiliary switch are on, the motor is on.}
\]

\[
\text{If the mainswitch is on the motor is on, or, if the auxiliary switch is on, the motor is on.}
\]

It has proved extremely difficult to give a pragmatic explanation of what is going wrong here.

Still, I think that it should be possible to draw a neat line between semantics and pragmatics, and not leave gaps like these. Even better, I think that the semantic theory developed in part III of this dissertation draws such a line, and that this line is drawn exactly as Grice's theory of conversation prescribes: Every counterexample to an argument form dubbed logically valid is to be explained as a product of a violation of the conversational maxims. (Note that Lewis is giving such an explanation). And every argument form
dubbed merely pragmatically correct, must have instantiations which show that the conclusion is sometimes cancellable. (Note that Adams does not give such an example.)

I.2.3. Logical form

The following quotation is taken from Geach (1976: 89).

'Roughly speaking, hypotheticals are sentences joined together with an 'if'. We don't count, however, sentences like I paid you back that fiver, if you remember; There's whisky in the decanter if you want a drink; for here the speaker is committed to asserting outright - not if something else is so - I paid you back that fiver or There's whisky in the decanter. Nor do we count sentences where 'if' means 'whether': I doubt if he'll come (quite good English, whatever nagging schoolmasters say). Nor do we count cases where 'if' has to be paraphrased with 'and': If you say that, he may hit you = Possibly you'll say that and he'll hit you; If it rains it sometimes thunders = Sometimes it rains and it thunders.'

Ever since 1905 and the publication of Russell's 'On denoting', the sort of distinction which Geach is implicitly making here between grammatical form on the one hand and logical form on the other has been quite familiar. For a while, during the heady youth of analytical philosophy, it even looked as if elucidating the logical form behind various misleading kinds of expressions would turn out to be the proper task of philosophy. It was a time in which one was largely interested in weeding out philosophy, in showing that large parts of traditional philosophy are in fact meaningless, a time in which Carnap could hope to show that Heidegger's work is nonsense just because it cannot be properly translated into standard predicate logical form.

These days we know that there is a lot more besides Heidegger which could not be translated into predicate logic, and not all of it is nonsense. Be this as it may, the notion of logical form is still very much with us, albeit in a modified role.
We have already mentioned the way a theory of pragmatics can and should complement a semantic theory, such that the two together cover as much of the whole field of what might be called intuitively sound argument forms as possible. In practice this does of course not always work out that well; an example of this is to be found in Adams' treatment of the Hypothetical Syllogism: \( \phi \rightarrow \psi, \psi \rightarrow \chi \rightarrow \phi \rightarrow \chi \). According to his logical theory it is not valid, while he cannot think of a pragmatic reason why arguments of this form in many cases seem acceptable. The way he then tacitly invokes the notion of logical form in order to reformulate the premises so that the argument becomes valid is typical of the role which it is all too often given in the literature on conditionals. (See Adams, 1975: 22).

"... we suggest that the 'hypothesis' of the first premise (the antecedent of the conditional) is tacitly 'presupposed' in the second, ... we will not attempt a rigorous justification of the foregoing intuitively plausible suggestion, but we will now see that if the suggestion is correct it would explain why apparent Hypothetical Syllogism inferences are rational, ... making the tacit presupposition of the second premise of a real life like Hypothetical Syllogism explicit, transforms it into an instance of the Restricted Hypothetical Syllogism pattern \( \phi \rightarrow \psi, (\psi \land \phi) \rightarrow \chi / \phi \rightarrow \chi \) which is universally probabilistically sound ..."

At its worst this is a strategy which cannot but result in a proliferation of epicycles, and Adams is by no means the only one. Cooper (1978:199) is embarrassed that his theory deems \( \sim(\phi \land \psi) \) and \( (\phi \rightarrow \sim\psi) \) equivalent:

"It seems reasonable to challenge

It is not the case that if Jones' car is gone he is out

If Jones' car is gone he is not out

It is unclear (to me) just what is going on in examples like these. Perhaps It is not the case that when followed by a conditional statement is sometimes understood to mean It is not necessarily the case that. Or perhaps negations of whole conditional statements, being rare in English, have an
interpretation which is idiosyncratic and simply unsettled at this point in the evolution of the language."

Recall that a similar jump from a negated conditional to an underlying logical form involving necessary was suggested in the first discussion of the marbles puzzle. It is a suggestion to which I am quite partial. As a matter of fact I wonder whether there are any negated English conditionals without this so-called hidden operator, but more about that presently. What I object to here is the ad-hoc manner in which such a possibility is introduced in order to save faces. If there are hidden operators in the negations of conditionals then the theory should be saying this. We shouldn't be forced to say it for the theory just to keep it sounding plausible.

A third example of the notion of logical form being used to patch up logical theories is to be found in the extensive literature on the argument form

\[(\varphi \lor \psi) \rightarrow \chi \not\rightarrow \varphi \rightarrow \chi,\]

which is invalid in most treatments of counterfactuals. This is usually thought unsatisfactory, but it turns out that you can render this argument form valid only at a fairly heavy price. If your theory says that logically equivalent sentences are interchangeable and that any sentence \(\varphi\) is logically equivalent to \((\varphi \land \psi) \lor (\varphi \land \neg \psi)\), then the validity of (*) implies the validity of

\[(**\) \ \varphi \lor \chi \not\rightarrow (\varphi \land \psi) \rightarrow \chi,\]

which in the context of counterfactuals is a lot less attractive. (Write \((\varphi \land \psi) \lor (\varphi \land \neg \psi)\) for its equivalent \(\varphi\) in \(\varphi \rightarrow \chi\) and then apply the principle in question once).

The reactions to this have been both many and varied. To be brief: Nute (1978, 1980) started by giving up the replaceability of logical equivalents, but later (1980 a) changed his position and defended (*) as pragmatically correct rather than logically valid. McKay and van Inwagen (1977) invented some counterexamples to (*), but most people think these are just pragmatically incorrect instances. Warmbröd (1981) accepts (**) and he finds its counterexamples pragmatically incorrect. But perhaps the
most common response (see Loewer (1978), Ellis (1979)) is just to deny that natural language counterfactuals with disjunctive antecedents can properly be formalized as \((\varphi \lor \psi) \rightarrow \chi\). Such a counterfactual is held to have the logical form \((\varphi \rightarrow \chi) \land (\psi \rightarrow \chi)\), and then the rest is easy enough.

In the following I will try my utmost not to use arguments like the above. Only if it is clear that dealing with an ambiguous sentence like for example

\textit{it is not the case that the red marble is in box 1 if the yellow marble is in box 2}

will I allow different logical forms: \(\sim(\text{yellow in 2} \rightarrow \text{red in 1})\) and \(\text{yellow in 2} \rightarrow \sim\text{red in 1}\). But I cannot see any ambiguity in Geach's example: \textit{If you say that, he may hit you}, therefore I will formalize it as a sentence of the form \((\varphi \rightarrow \text{may } \psi)\) — that is, as closely to the surface structure as possible. For the same reason I will formalize English sentences which run like 'it is not the case that if ..., then ...' as \(\sim(\ldots \rightarrow \ldots)\). If you were to say these English sentences say that it is not necessarily the case that if ... then ..., then I would agree. But if you were to conclude from that that they should be formalized as \(\sim \circ (\ldots \rightarrow \ldots)\), then at least you should be able to come up with an example of a negated conditional sentence which does not say that it is not necessarily so that if ... then ...

The reason why I follow this strategy is because I think it is the most sensible one to follow in studies like these, where hardly any attention is paid to syntactic questions. It is not because I think that there is no need for a 'level of logical form' in syntax, or that such a level is only needed for desambiguation, (as it is in Montague Grammar) I am ready to admit that syntax is just as important to semantics as pragmatics is, and that ultimately the question is which combined theory of syntax + semantics + pragmatics offers the best explanations. It might be that such a theory will enable us to use syntactic arguments in our explanations. It might even be that the notion of logical form will then be
of key importance. In the absence of such a theory, however, it is best to avoid ad-hoc explanations.
NOTES TO PART I

1. ~ is short for not, \( \land \) is short for and, \( \lor \) is short for or, and \( \rightarrow \) is short for if \( \ldots \) then \( \ldots \).

2. This example is a slight variant of Lewis Carroll's Barbershop Paradox, which first appeared in Carroll (1894).

3. Modus Tollens does fail in the theory of conditionals put forward in Cooper (1978). For this particular example, however, it holds.

4. So wrote Theodor Lipps in 1880. (Quoted by Chisholm (1966: 79)).

5. I do not think that the conventionalist is entitled to draw this conclusion. In doing so, he is ignoring the possibility that sometimes language users cannot follow the conventions governing the use of a given logical constant without certain ontological assumptions being made.

6. For a recent defense of this position see Ellis (1979).

7. I think that the first to conceive of this possibility were the Neo-Kantians.

8. See for example the first chapter of Nute (1980).

9. There is a ready analogy with the way the Chomsky tradition sees syntactic competence.

10. My own views on the status of logic have been greatly influenced by the teachings of Else Barth. See for example
Barth (1979).

11. This has been stressed more than once by Michael Dummett.

12. This does not mean, however, that I am convinced that for our purposes this really is the best way to proceed. I am aware that if one takes a constructive stand on the ontological status of mathematical objects - as perhaps one should - one just cannot talk the way classical logic suggests. Then, the way of speaking underlying intuitionistic logic is much more suitable.

13. For other ways see II.2.2.

14. C.f. the discussion of possible worlds in chapter II.1.1.

15. A similar result is proved by Gibbard (1981).

16. This could have been done otherwise, but for the present purposes it would not make any difference.


18. Harper et. al. (1981) contains most of the important papers on the subject. (This anthology is mentioned in the references under Gibbard (1981).)

19. For a detailed criticism of relevance logic see Copeland (1978).

20. A working knowledge of his work is assumed here.


22. See Groenendijk and Stokhof (1982) for further discussion of these points.
23. Perhaps Discourse Representation Theory (see Kamp (1984)) can be seen as a new theory of logical form.