Logical Consequence

When we say statement A is true ‘because of’ statement B, there can be a variety of links between the statements. B might describe the physical cause of A, or the real situation which makes it true. Some links, though, don’t seem to depend on how the world is, but on the contents or form of A and B, and ‘logical consequence’ is a link of this kind. An understanding of this link could reveal the nature of reason itself, and hence it is a key issue in modern philosophy.

As so often in analytic philosophy, one option is to treat logical consequence as ‘primitive’, meaning we all know what it is, but it can’t be explained, so we should just accept it. That is a last resort, and a slightly less pessimistic view is to treat consequence as a ‘black box’, meaning you are ignorant of its deep nature, but can map its behaviour thoroughly and accurately. A standard approach of this type is to map the inputs and outputs in terms of their truth or falsity, leading to the sweeping proposal that we have logical consequence whenever true inputs always lead to true outputs, so that consequence only fails if a group of truths can lead to a falsehood.

The details of this proposal are worked out in ‘model theory’. The key first step is to say that there is a sharp distinction between the vocabulary of a sentence which is ‘logical’ (connectives like ‘and’, or implication words like ‘so’), and ‘non-logical’ (such as words for objects and predicates). In natural language this may be a little blurred, but in symbolic languages the distinction can be precise. A model is a set of sentences in which the non-logical vocabulary is replaced by variables, and then each model can have ‘interpretations’, which are complete replacements of all the variables by selections from sets of objects and predicates. This means that the logical vocabulary fixes the logical framework of a particular model, and a large number of interpretations of the model can then be specified. The last step is to say that we have logical consequence when a model of an argument retains a true output for all true interpretations, and that tells you all you can know about logical consequence. The account is backed up by a formal theory of the nature of truth, and is sometimes interpreted in terms of ‘possible worlds’ where models are true. A further suggestion is to test consequence by seeing if it survives varying the real world situation, rather than the interpretation of the variables. A formally valid argument should survive either test.

This account is the modern orthodox view, because it fits the role of logical validity in our thinking, and it is precise, and fits into many other formal accounts of philosophical issues in analytic philosophy. There are, nevertheless, discontents, because of its ‘black box’ character – that is, it maps consequence, but doesn’t explain it. Two phenomena, in particular, are not captured by this model-theoretic account of consequence. The first is that it says nothing about the reason *why* some set of truths requires that some further truth should hold, and the second is that it doesn’t capture the feeling that a logical consequence *must* entail some truth.

The first discontent is shown if you make the consequence of your model a statement which is always true, such as ‘…so wombats are animals’, and then construct a model that says nothing about wombats, since this will qualify as a logical consequence in the model-theoretic account. It has not explained our ordinary notion of logical consequence.

A first step towards explaining why some logical consequence holds is to distinguish between consequence by proof, and consequence by meaning. In a purely formal language, expressed entirely in symbols, such as predicate calculus, it is possible to prove many things, without ever interpreting what the variables mean. Hence we have a concept of logical consequence which doesn’t involve interpretation, and this is called ‘proof-theoretic’ consequence, written ‘⊢’. If we write ‘Γ ⊢ φ’, this reads as ‘the set of formulas labelled *Gamma* always proves the formula labelled *phi’*. This is a very formal view of consequence.

On the other hand, if we say ‘it is red so it is coloured’ or ‘it is square so it can’t be circular’, we accept their validity, but there is no proof (so maybe there are rational consequences which are not strictly ‘logical’). For this there is the symbol ‘ ⊨’, and ‘Γ ⊨ φ’ can be read as ‘*phi* is a consequence of the meaning or concepts of *gamma’.* There is some ambiguity about this second symbol, because it also used for the model-theoretic concept of consequence outlined above, and ‘ ⊨’ can be read as ‘models’. This is because models involve truth and falsehood of formulas, and so count as ‘semantic’. A consequence is said to be ‘material’ if it involves content as well as form. We can disprove a case of ‘⊢’ with a formal proof, but disproving ‘ ⊨’ needs a counterexample, such as something which is red but not coloured, or some false consequence of an apparently true model.

One response to this distinction is to embrace ‘pluralism’ about logical consequence. If there is proof-theoretic and semantic consequence, and also a different concept of consequence for each non-classical logic, then the hunt for the true ‘nature’ of consequence may be in vain. ‘Logical consequence’ may be as vague as ‘because of’ in ordinary talk.

The second discontent is that there is a traditional view (still widely accepted) that consequence is a stronger relation than the description offered by model theory, because the premises should *necessitate* their valid conclusion, suggesting a more active connection across the link. This view, though, requires some account of what does the necessitating, and that is not obvious. It seems unlikely, for example, that the force of necessity just arises from the fixed conventions of formal logic, since the same necessity is felt in ordinary informal reasoning. One possibility is that it only counts as ‘logical’ consequence if longs chains of reasoning can be sustained (that is, that consequence is ‘transitive’). If confidence in your chain of reasoning gradually dwindles, that shows that you are not doing strict logic. It might be that it is the unshakable nature of reality which gives us this feeling of necessity in an argument, suggesting that logical consequence is rooted in the physical world, rather than in pure reason. A rival view might say that the necessity comes from a direct a priori grasp of the structure of the unchanging ‘space of reasons’.

The authority of logical consequence is captured in the *modus ponens* rule, which says that if you accept some statement, you must also accept its known logical consequences. The concept of ‘material implication’ treats ‘A→B’ as just saying that B cannot be false if A is true, but this resembles model-theoretic consequence, and similarly doesn’t seem to quite capture what we mean by ‘conditionals’ (truths expressed by ‘if A, then B’).