1. Accessing S2

Here is one problem that Lewis identifies: There will be sentences, T, meeting the following conditions:

(1)  
   a. T is a theorem of S2
   b. There are sentences \( \varphi \) and \( \psi \) such that \( (T \land \varphi) \Rightarrow \psi \) is a theorem of S2
   c. But \( \varphi \Rightarrow \psi \) is not a theorem of S2.

In other word, the fact that both T and \( (T \land \varphi) \Rightarrow \psi \) are theorems of S2 does not guarantee that \( \varphi \Rightarrow \psi \) is a theorem of S2. This is a problem, since it violates our intuition about logical reasoning. Our intuition is that once we have established something as a theorem, we should be able to presuppose its truth. What's the point of theorems otherwise? If \( \psi \) is deducible from \( \varphi \) with the help of a theorem T, then that should mean that it is deducible from \( \varphi \) alone. So there seems to be something incomplete about S2. The same property carries over to S3. It disappears only when you get to S4. In fact, it is absent in any "normal system".

Here is another fact about S2.

(2)  
   a. \( \models_{S2} \lozenge(p \lor \Box(q \lor \neg q)) \)
   b. but neither \( \models_{S2} \lozenge p \)
   c. nor \( \models_{S2} \lozenge \Box(q \lor \neg q) \)

We have a disjunction which is a theorem, neither of whose disjunct is a theorem. Moreover, there is no sentence letter in common between the two disjuncts. The worry here is that if \( \lozenge \lozenge p \) is not a theorem, then in an appropriate semantics, there must be a way – a valuation – to make it false. Similarly, there must be a valuation to make \( \Box(q \lor \neg q) \) false, since it is not a theorem either. Now the question becomes why can we not put those two valuations together to make \( \lozenge \lozenge p \lor \Box(q \lor \neg q) \) false. Either the proof system is too weak to make one of the disjuncts a theorem, or it is too strong and makes the disjunction a theorem, when it can in fact be falsified.

These are two signs that S2 is not satisfactory. Either we have to add something to it, or take something away from it.

2. Finding a semantics Modal Logic

2.1. The metalinguistic strategy

The metalinguistic approach takes '\( \Box \)' to be a predicate of sentences as syntactic objects. Let's start with this.
(3)  ⌈□φ⌉ is true iff |- φ

This is intended to mean that φ is necessary iff φ is a theorem, i.e. a logical truth. Now (3) actually says something about sentences. It says that prefixing a formula – any formula – with a box will produce a sentence which is true iff what follows the box in that sentence is a theorem, i.e. is derivable from the axioms of the system. It then follows that '□p' is true iff p is a theorem. But p is a sentence letter. The theoremhood of p means the relevant system is inconsistent: the substitution rule allows us to replace p with any other formula, and a system is inconsistent iff every formula is a theorem of it. Thus, (3) seems to imply that '□p' is true iff the system is inconsistent. But our starting point is that the relevant system is consistent. It can indeed be shown that S1, S2, S3, S4 are all consistent. Thus, (3) means '□p' is false. As the choice of p is arbitrary, saying '□p' is false amounts to saying that '□p' is inconsistent, which means ¬□p is a theorem, which means ◊¬p is a theorem, which means any formula is a theorem (since p can be substituted by any formula), which means the system is inconsistent, which contradicts our initial assumption. Thus, (3) does not work.

We have a clear notion of tautology: a Propositional Logic (PL) formula which is logically true, i.e. which is true no matter what values are given to the sentence letters in it. So maybe we can use the notion of tautology to get to a semantics for Modal Logic. Here is an attempt.

(4)  ⌈□φ⌉ is true iff φ is a tautology

What this says is that substituting a tautology for ⌈φ⌉ in ⌈□φ⌉ will get us a true sentence. The problem is this. Take a formula like ¬p. Prefixing it with a square gets us □¬p. By definition, □¬p is not a tautology: it is not a sentence of PL. So we can deduce from (4) that □¬p is false, no matter what value p has. This means that ¬□¬p is true for any value of p, which in turns means that ◊◊p is a theorem. But this is clearly not true. It can be shown that ◊◊p is not a theorem of S2. Thus, (4) doesn't work!

We could require that ⌈□φ⌉ is true iff φ is a logical truth – not of PL – but of Modal Logic itself - the system for which we are developing a semantics. But this leads to a circularity, since we are developing a semantics for Modal Logic by appealing to a notion – "logical truth of Modal Logic" – whose meaning depends on the existence of such a semantics.

However, we could use the notion "logical truth of Modal Logic" to constrain the relevant system without knowing exactly what it means. Specifically, we could require that whatever is a logical truth of Modal Logic, prefixing it with a square will result in a logical truth of Modal Logic. In other word,

(5)  if |- φ, then |- □φ

This is the necessitation rule which distinguishes a "normal system" from a non-normal system. A system which has it will have □□(q v ¬q) as a theorem. In fact,
adding (5) to the rules of S2 – or equivalently – adding □(q ∨ ¬q) to its axioms – will give us the system T, which is weaker than S4 but stronger than S2.

So we can ask whether T is the "right" system of Modal Logic, i.e. a system that understands the box to mean logical truth in the system itself. The answer is "we can't tell", since (5) still does not say what the meaning of the box is.

**2.2. The truth-table strategy**

Let’s suppose that the box is a predicate – not of sentences – but of what they express, i.e. propositions. What are then propositions? This is a rather involved philosophical question. We need not answer it here. Instead, we can try to identify a property of propositions and develop the semantics for '□' based on that property. We know that propositions have truth-values. Would truth-values be enough? The answer seems to be no, since it is rather obvious that '□' is not truth-functional, as becomes evident when we attempt to give a truth-table for it.

<table>
<thead>
<tr>
<th>φ</th>
<th>□φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1

When φ is false, then □φ is definitely false. But when φ is true, what should the truth-value of □φ be? If it is "true", then we trivialize Modal Logic: it would be Propositional Logic with the addition of a symbol which doesn't do anything. If it is "false", □φ becomes a contradiction, certainly not how we understand "necessity" intuitively.

We thus need a more fine-grained distinction. Suppose that besides truth-value, propositions also have modal status: a proposition is not only true or false, it is also necessary or contingent. The distinction then becomes four-way.

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Necessary</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Contingent</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2

A necessary truth has the value 1, a contingent truth the value 2 etc. Here is the truth-table for the two one-place operators '□' and '¬'.

<table>
<thead>
<tr>
<th>φ</th>
<th>¬φ</th>
<th>□φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 3

This seems intuitively satisfactory. Negation changes truth-value without changing modal status. Necessitation of a necessary truth (1) should indeed yield a necessary
truth: \(\Box(1) = 1\). Saying of something contingently true that it is necessarily true should indeed be necessarily false: \(\Box(2) = 4\). And if something is false – either contingently or necessarily – it certainly cannot be the case that it is necessarily true: \(\Box(3) = \Box(4) = 4\).

So our problem is solved! Well not yet. Let's consider conjunction.

From this table, we see that conjoining anything with a necessary truth will not change its value, and conjoining anything with a necessary falsehood will yield a necessary falsehood. This corresponds to our intuition. So the four sides of the square should be as they are: two rows of 1-2-3-4 and two rows of 4-4-4-4. Furthermore, conjunction should be idempotent, so the north-west south-east diagonal should be as it is: 1-2-3-4. Now what about the two remaining cells.

These cells represent the conjunction of a contingent truth (2) with a contingent falsehood (3). What should the result be? Let the contingent truth be "Obama is president" and the contingent falsehood be "Sean Penn is vice-president", then obviously their conjunction will be a contingent falsehood (3). But suppose the contingent falsehood is "Obama is not president". Then the result, obviously, will be a necessary falsehood (4). So these cells can be either 3 or 4!

This is clearly not a satisfactory solution. The lesson is: when we make things finer, we gain truth-functionality in one place but loose it in another place.